

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2024-502028

(P2024-502028A)

(43)公表日 令和6年1月17日(2024.1.17)

(51)国際特許分類	F I	テーマコード (参考)
C 1 2 N 15/11 (2006.01)	C 1 2 N 15/11	Z Z N A 4 B 0 6 3
C 1 2 N 9/10 (2006.01)	C 1 2 N 9/10	4 B 0 6 5
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869	Z
C 1 2 N 9/16 (2006.01)	C 1 2 N 9/16	Z
C 1 2 N 5/071(2010.01)	C 1 2 N 5/071	
審査請求 未請求 予備審査請求 未請求 (全171頁)		

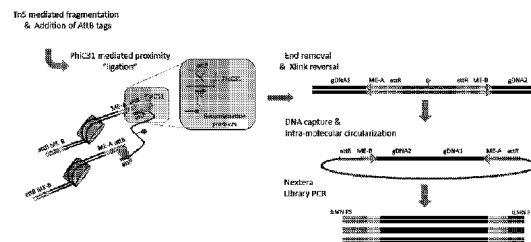
(21)出願番号	特願2023-539958(P2023-539958)	(71)出願人	517286593
(86)(22)出願日	令和3年12月29日(2021.12.29)		ダブテイル ゲノミクス エルエルシー
(85)翻訳文提出日	令和5年8月29日(2023.8.29)		アメリカ合衆国 9 5 0 6 6 カリフォル
(86)国際出願番号	PCT/US2021/065519		ニア州 スコッツバレー エンタープライ
(87)国際公開番号	WO2022/147129		ズ・ウェイ 1 0 0 スイート エー1 0 1
(87)国際公開日	令和4年7月7日(2022.7.7)	(74)代理人	110003797
(31)優先権主張番号	63/132,304		弁理士法人清原国際特許事務所
(32)優先日	令和2年12月30日(2020.12.30)	(72)発明者	ムンディング, エリザベス
(33)優先権主張国・地域又は機関	米国(US)		アメリカ合衆国 9 5 0 6 6 カリフォル
(31)優先権主張番号	63/145,859		ニア州 スコッツバレー エンタープライ
(32)優先日	令和3年2月4日(2021.2.4)	(72)発明者	ズ・ウェイ 1 0 0 スイート エー1 0 1
(33)優先権主張国・地域又は機関	米国(US)		ブランシェット, マルコ
(31)優先権主張番号	63/216,170		アメリカ合衆国 9 5 0 6 6 カリフォル
	最終頁に続く		ニア州 スコッツバレー エンタープライ
			ズ・ウェイ 1 0 0 スイート エー1 0 1
			最終頁に続く

(54)【発明の名称】 配列決定ライブラリー調製のための方法および組成物

(57)【要約】

【解決手段】本明細書では、近接ライゲーションの方法およびそのような方法で使用するための組成物が提供される。単一細胞核酸立体構造の評価、または単一細胞核酸配列もしくはフェーシング情報の決定に関する実施形態も本明細書で提供される。立体構造が保存されたまたは立体構造が再構築された核酸試料は、試料から生成されたペアエンドライブラリーの分析時に、ペアエンドがパーティション起源または細胞起源に割り当てられるように、アリコート識別配列セグメントが付加されたアリコートに断片化および分布させることができる。したがって、配列と3次元核酸配置における細胞特異的な変化を判定することができる。

【選択図】図3 2



【特許請求の範囲】**【請求項 1】**

核酸処理の方法であって、前記方法は、

(a) 少なくとも 1 つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程と、

(b) 前記核酸分子を、少なくとも第 1 のセグメントおよび第 2 のセグメントを含む複数のセグメントへ切断する工程であって、切断する工程はトランスポザーゼによって達成される、工程と、

(c) 前記第 1 のセグメントおよび前記第 2 のセグメントに第 1 のリコンビナーゼ部位を結合させる工程と、

(d) リコンビナーゼの存在下で、第 2 のリコンビナーゼ部位を含むリンカーに、前記第 1 のセグメントおよび前記第 2 のセグメントを接触させる工程であって、それによって、前記第 1 のセグメント由来の第 1 の配列と、前記リンカー由来のリンカー配列と、前記第 2 のセグメント由来の第 2 の配列とを含む結合された核酸を生成する、工程と、
を含む、方法。

10

【請求項 2】

前記トランスポザーゼは T n 5 トランスポザーゼである、請求項 1 に記載の方法。

【請求項 3】

前記第 2 のリコンビナーゼ部位は、直接反復配列として配向された 2 つのリコンビナーゼ部位を含む、請求項 1 に記載の方法。

20

【請求項 4】

前記リコンビナーゼはインテグラーゼである、請求項 1 に記載の方法。

【請求項 5】

前記インテグラーゼは、P h i C 3 1 インテグラーゼ、B x b 1 インテグラーゼ、またはそれらの組み合わせである、請求項 4 に記載の方法。

【請求項 6】

前記結合された核酸の 5' 末端を、前記結合された核酸の 3' 末端にライゲーシオンすることによって、前記結合された核酸を環状化する工程をさらに含む、請求項 1 に記載の方法。

【請求項 7】

前記結合された核酸の少なくとも一部を配列決定する工程をさらに含む、請求項 1 に記載の方法。

30

【請求項 8】

前記配列決定する工程は、前記第 1 の配列の少なくとも一部および前記第 2 の配列の少なくとも一部を配列決定することを含む、請求項 7 に記載の方法。

【請求項 9】

前記第 1 の配列の少なくとも一部および前記第 2 の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む、請求項 8 に記載の方法。

【請求項 10】

前記配列決定する工程からの情報を使用して、3次元ゲノムの分析を行う工程をさらに含む、請求項 7 に記載の方法。

40

【請求項 11】

前記安定化された試料はクロスリンクした試料である、請求項 1 に記載の方法。

【請求項 12】

前記安定化された試料を得る工程は、試料を得ることと、前記試料を安定化することとを含む、請求項 1 に記載の方法。

【請求項 13】

前記安定化された試料を得る工程は、前もって安定化された試料を得ることを含む、請求項 1 に記載の方法。

【請求項 14】

50

前記核酸結合タンパク質は、クロマチンまたはその成分を含む、請求項 1 に記載の方法。

【請求項 1 5】

前記第 1 のリコンビナーゼ部位および前記第 2 のリコンビナーゼ部位は、a t t P および a t t B インテグラーゼ部位を含む、請求項 1 に記載の方法。

【請求項 1 6】

前記第 1 のリコンビナーゼ部位は、前記第 2 のリコンビナーゼ部位とは異なる、請求項 1 に記載の方法。

【請求項 1 7】

前記第 1 のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位である、請求項 1 に記載の方法。 10

【請求項 1 8】

前記第 2 のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位である、請求項 1 に記載の方法。

【請求項 1 9】

前記結合された核酸は、ヘアピンループを形成しない、請求項 1 に記載の方法。

【請求項 2 0】

前記第 1 のリコンビナーゼ部位および前記第 2 のリコンビナーゼ部位は、トランスポザーゼモザイク末端を含む、請求項 1 に記載の方法。

【請求項 2 1】

前記リンカー配列はバーコード配列を含む、請求項 1 に記載の方法。 20

【請求項 2 2】

前記バーコード配列はパーティション起源を示す、請求項 2 1 に記載の方法。

【請求項 2 3】

前記バーコード配列は細胞起源を示す、請求項 2 1 に記載の方法。

【請求項 2 4】

前記バーコード配列は細胞集団起源を示す、請求項 2 1 に記載の方法。

【請求項 2 5】

前記バーコード配列は生物起源を示す、請求項 2 1 に記載の方法。

【請求項 2 6】

前記バーコード配列は種起源を示す、請求項 2 1 に記載の方法。 30

【請求項 2 7】

前記方法は 4 時間以下で完了する、請求項 1 に記載の方法。

【請求項 2 8】

前記安定化された試料は、50,000 個以下の細胞を含む、請求項 1 に記載の方法。

【請求項 2 9】

前記安定化された試料は、少なくとも 10,000 個の細胞を含む、請求項 2 8 に記載の方法。

【請求項 3 0】

前記安定化された試料、安定化された核を含む、請求項 1 に記載の方法。 40

【請求項 3 1】

前記安定化された試料は、50,000 個以下の核を含む、請求項 1 に記載の方法。

【請求項 3 2】

前記安定化された試料は、少なくとも 10,000 個の核を含む、請求項 3 1 に記載の方法。

【請求項 3 3】

前記安定化された試料は超音波処理されない、請求項 1 ～ 3 2 のいずれか 1 つに記載の方法。

【請求項 3 4】

核酸処理の方法であって、前記方法は、 50

(a) 少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程と、

(b) 前記核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントへ切断する工程と、

(c) 前記第1のセグメントおよび前記第2のセグメントに第1のリコンビナーゼ部位を結合させる工程と、

(d) リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、前記第1のセグメントおよび前記第2のセグメントを接触させる工程であって、それによって、前記第1のセグメント由来の第1の配列と、前記リンカー由来のリンカー配列と、前記第2のセグメント由来の第2の配列とを含む結合された核酸を生成し、前記第2のリコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む、工程と、

10

を含む、方法。

【請求項35】

前記切断する工程はトランスポザーゼによって達成される、請求項34に記載の方法。

【請求項36】

前記トランスポザーゼはTn5トランスポザーゼである、請求項35に記載の方法。

【請求項37】

前記リコンビナーゼはインテグラーゼである、請求項34に記載の方法。

【請求項38】

20

前記インテグラーゼは、PhiC31インテグラーゼ、Bxb1インテグラーゼ、またはそれらの組み合わせである、請求項34に記載の方法。

【請求項39】

前記結合された核酸の5'末端を、前記結合された核酸の3'末端にライゲーションすることによって、前記結合された核酸を環状化する工程をさらに含む、請求項34に記載の方法。

【請求項40】

前記結合された核酸の少なくとも一部を配列決定する工程をさらに含む、請求項34に記載の方法。

【請求項41】

30

前記配列決定する工程は、前記第1の配列の少なくとも一部および前記第2の配列の少なくとも一部を配列決定することを含む、請求項38に記載の方法。

【請求項42】

前記第1の配列の少なくとも一部および前記第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む、請求項41に記載の方法。

【請求項43】

前記配列決定する工程からの情報を使用して、3次元ゲノムの分析を行う工程をさらに含む、請求項34に記載の方法。

【請求項44】

前記安定化された試料は、クロスリンクした試料である、請求項34に記載の方法。

40

【請求項45】

前記安定化された試料を得る工程は、試料を得ることと、前記試料を安定化することとを含む、請求項34に記載の方法。

【請求項46】

前記安定化された試料を得る工程は、前もって安定化された試料を得ることを含む、請求項34に記載の方法。

【請求項47】

前記核酸結合タンパク質は、クロマチンまたはその成分を含む、請求項34に記載の方法。

【請求項48】

50

前記第1のリコンビナーゼ部位および前記第2のリコンビナーゼ部位は、a t t Pおよびa t t Bインテグラーゼ部位を含む、請求項34に記載の方法。

【請求項49】

前記第1のリコンビナーゼ部位は、前記第2のリコンビナーゼ部位とは異なる、請求項34に記載の方法。

【請求項50】

前記第1のリコンビナーゼ部位は、a t t Pまたはa t t Bインテグラーゼ部位である、請求項34に記載の方法。

【請求項51】

前記第2のリコンビナーゼ部位は、a t t Pまたはa t t Bインテグラーゼ部位である、請求項34に記載の方法。 10

【請求項52】

前記結合された核酸は、ヘアピンループを形成しない、請求項34に記載の方法。

【請求項53】

前記第1のリコンビナーゼ部位および前記第2のリコンビナーゼ部位は、トランスポザーゼモザイク末端を含む、請求項34に記載の方法。

【請求項54】

前記リンカー配列はバーコード配列を含む、請求項34に記載の方法。

【請求項55】

前記バーコード配列は、パーティション起源を示す、請求項54に記載の方法。 20

【請求項56】

前記バーコード配列は、細胞起源を示す、請求項54に記載の方法。

【請求項57】

前記バーコード配列は、細胞集団起源を示す、請求項54に記載の方法。

【請求項58】

前記バーコード配列は、生物起源を示す、請求項54に記載の方法。

【請求項59】

前記バーコード配列は、種起源を示す、請求項54に記載の方法。

【請求項60】

前記方法は4時間以下で完了する、請求項34に記載の方法。 30

【請求項61】

前記安定化された試料は、50,000個以下の細胞を含む、請求項34に記載の方法。

【請求項62】

前記安定化された試料は、少なくとも10,000個の細胞を含む、請求項61に記載の方法。

【請求項63】

前記安定化された試料は、安定化された核を含む、請求項34に記載の方法。

【請求項64】

前記安定化された試料は、50,000個以下の核を含む、請求項34に記載の方法。 40

【請求項65】

前記安定化された試料は、少なくとも10,000個の核を含む、請求項64に記載の方法。

【請求項66】

前記安定化された試料は超音波処理されない、請求項34～65のいずれか1つに記載の方法。

【請求項67】

方法であって、

(a) 少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、

(b) 複合体を形成するために前記核酸分子をデンドリマーと接触させる工程であって、デンドリマーの1つ以上のポリマーが末端の一級アミンを含む、工程と、

(c) 前記核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントへ切断する工程と、

(d) 前記複数のセグメントの第1のセグメントおよび第2のセグメントを接合部で結合させる工程と、を含む、方法。

【請求項68】

前記デンドリマーはクロスリンカーで修飾される、請求項67に記載の方法。

【請求項69】

(b)の前に、前記デンドリマーをクロスリンカーに接触させる工程をさらに含む、請求項67に記載の方法。

【請求項70】

前記クロスリンカーはソラレン、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファミド(isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン(ciccy cloplatin)、エプタプラチン(eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド(mitozolomide)、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキシソルピシン、ダウノルピシン、エビルピシン、またはイダルピシンを含む、請求項68または69に記載の方法。

【請求項71】

前記ソラレンはN-ヒドロキシスクシンイミド(NHS)エステル-コンジュゲートソラレンを含む、請求項70に記載の方法。

【請求項72】

前記デンドリマーはポリアミドアミン(PAMAM)デンドリマーを含む、請求項67~71のいずれか1つに記載の方法。

【請求項73】

(e) 前記デンドリマーから前記クロスリンカーを分離する工程をさらに含む、請求項68~72のいずれか1つに記載の方法。

【請求項74】

分離する工程は熱アルカリ処理を含む、請求項73に記載の方法。

【請求項75】

分離する工程はUV放射への曝露を含む、請求項73に記載の方法。

【請求項76】

前記複数のセグメントの一部は結合してコンカテマーを形成する、請求項67~75のいずれか1つに記載の方法。

【請求項77】

前記コンカテマーは少なくとも3個のセグメントを含む、請求項76に記載の方法。

【請求項78】

前記コンカテマーは少なくとも4個のセグメントを含む、請求項76に記載の方法。

【請求項79】

前記コンカテマーは少なくとも5個のセグメントを含む、請求項76に記載の方法。

【請求項80】

前記コンカテマーは少なくとも6個のセグメントを含む、請求項76に記載の方法。

【請求項81】

前記コンカテマーは少なくとも8個のセグメントを含む、請求項76に記載の方法。

【請求項82】

10

20

30

40

50

前記コンカテマーは少なくとも10個のセグメントを含む、請求項76に記載の方法。

【請求項83】

前記デンドリマーの分子量は、5キロダルトン(kDa)～125kDaである、請求項67～82のいずれか1つに記載の方法。

【請求項84】

前記デンドリマーの分子量は、6kDa～8kDaである、請求項67～83のいずれか1つに記載の方法。

【請求項85】

前記デンドリマーの分子量は、25kDa～35kDaである、請求項67～83のいずれか1つに記載の方法。

【請求項86】

前記デンドリマーの分子量は、110kDa～125kDaである、請求項67～83のいずれか1つに記載の方法。

【請求項87】

前記デンドリマーは32～512個の反応基を含む、請求項67～82のいずれか1つに記載の方法。

【請求項88】

前記デンドリマーは約32個の反応基を含む、請求項67～87のいずれか1つに記載の方法。

【請求項89】

前記デンドリマーは約128個の反応基を含む、請求項67～87のいずれか1つに記載の方法。

【請求項90】

前記デンドリマーは約512個の反応基を含む、請求項67～87のいずれか1つに記載の方法。

【請求項91】

(b)の後に、デンドリマー複合体を光活性化する工程をさらに含む、請求項67～90のいずれか1つに記載の方法。

【請求項92】

(f)複数の選択されたセグメントを得るために、前記複数のセグメントをサイズ選択に供する工程をさらに含む、請求項67～91のいずれか1つに記載の方法。

【請求項93】

切断する工程は、前記核酸分子をデオキシリボヌクレアーゼ(DNase)と接触させることを含む、請求項67～92のいずれか1つに記載の方法。

【請求項94】

DNaseは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む、請求項93に記載の方法。

【請求項95】

前記安定化された生体試料はクロスリンク剤で処理される、請求項67～94のいずれか1つに記載の方法。

【請求項96】

前記クロスリンク剤は化学的固定剤である、請求項95に記載の方法。

【請求項97】

前記化学的固定剤は、ホルムアルデヒド、ソラレン、ジスクシンイミジルグルタレート(DSG)、エチレングリコールビス(スクシンイミジルスクシネート)(EGS)、紫外線、またはそれらの組み合わせを含む、請求項96に記載の方法。

【請求項98】

前記クロスリンク剤は、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ペンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファ

10

20

30

40

50

ミド (i s o f a m i d e) 、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン (c i c y c l o p l a t i n) 、エプタプラチン (e p t a p l a t i n) 、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (m i t o z o l o m i d e) 、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンを含む、請求項 9 5 に記載の方法。

【請求項 9 9】

前記安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である、請求項 6 7 ～ 9 4 のいずれか 1 つに記載の方法。

10

【請求項 1 0 0】

前記安定化された生体試料は安定化された細胞溶解物を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

【請求項 1 0 1】

前記安定化された生体試料は安定化された無傷細胞を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

【請求項 1 0 2】

前記安定化された生体試料は安定化された無傷の核を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

20

【請求項 1 0 3】

工程 (c) は、前記無傷細胞または前記無傷の核の溶解の前に実施される、請求項 1 0 1 または 1 0 2 に記載の方法。

【請求項 1 0 4】

工程 (d) の前に、前記安定化された生体試料中で細胞および／または核を溶解させる工程をさらに含む、請求項 6 7 ～ 1 0 0 のいずれか 1 つに記載の方法。

【請求項 1 0 5】

前記安定化された生体試料は 3 , 0 0 0 , 0 0 0 未満の細胞を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

【請求項 1 0 6】

前記安定化された生体試料は 1 , 0 0 0 , 0 0 0 未満の細胞を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

30

【請求項 1 0 7】

前記安定化された生体試料は 1 0 0 , 0 0 0 未満の細胞を含む、請求項 6 7 ～ 9 8 のいずれか 1 つに記載の方法。

【請求項 1 0 8】

結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を f i l l e r i n すること、および平滑末端をライゲーションすることを含む、請求項 6 7 ～ 1 0 7 のいずれか 1 つに記載の方法。

【請求項 1 0 9】

結合させる工程は、少なくとも 1 つの架橋オリゴヌクレオチドに、少なくとも前記第 1 のセグメントおよび前記第 2 のセグメントを接触させることを含む、請求項 6 7 ～ 1 0 7 のいずれか 1 つに記載の方法。

40

【請求項 1 1 0】

前記架橋オリゴヌクレオチドはバーコード配列を含む、請求項 1 0 9 に記載の方法。

【請求項 1 1 1】

結合させる工程は、少なくとも前記第 1 のセグメントおよび前記第 2 のセグメントを、複数の架橋オリゴヌクレオチドに連続して接触させることを含む、請求項 1 1 0 に記載の方法。

【請求項 1 1 2】

50

結合させる工程は、架橋オリゴヌクレオチドの固有の配列を受け取る安定化された生体試料の試料、細胞、核、染色体、または核酸分子をもたらす、請求項 109 に記載の方法。

【請求項 113】

結合させる工程は、少なくとも前記第 1 のセグメントおよび前記第 2 のセグメントをバーコードに接触させることを含む、請求項 67～107 のいずれか 1 つに記載の方法。

【請求項 114】

(g) 第 1 のリード対を生成するために前記接合部の各側で少なくともいくつかの配列を得る工程をさらに含む、請求項 67～113 のいずれか 1 つに記載の方法。

【請求項 115】

(h) 前記第 1 のリード対をコンティグのセットにマッピングする工程と、

(i) ゲノムに対する順序および／または配向を表す前記コンティグのセットを通る経路を決定する工程と、

をさらに含む、請求項 114 に記載の方法。

【請求項 116】

(h) 前記第 1 のリード対をコンティグのセットにマッピングする工程と、

(i) 前記コンティグのセットから、前記安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程と、

をさらに含む、請求項 114 に記載の方法。

【請求項 117】

(h) 前記第 1 のリード対をコンティグのセットにマッピングする工程と、

(i) 前記コンティグのセット内のバリエーションをフェーズに割り当てる工程と、

をさらに含む、請求項 114 に記載の方法。

【請求項 118】

(h) 前記第 1 のリード対をコンティグのセットにマッピングする工程と、

(i) 前記コンティグのセットから、前記コンティグのセットにおけるバリエーションの存在を決定する工程と、

(j) (1) 前記安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2) 前記バリエーションの存在に基づいて薬物を選択すること、または、

(3) 前記安定化された生体試料についての薬物効力を同定すること、のうちの 1 つ以上から選択される工程を行う工程と、

をさらに含む、請求項 114 に記載の方法。

【請求項 119】

分析の方法であって、前記方法は、

(a) 少なくとも 1 つの核酸結合タンパク質に複合体化された核酸分子および少なくとも 1 つの非コード RNA を含む安定化された生体試料を得る工程と、

(b) Tn5 トランスポザターゼ、ならびに、モザイク末端および検出可能な標識を含むオリゴヌクレオチドに、DNA 分子を接触させる工程であって、それによって前記 DNA 分子を断片化し、断片化した DNA 分子の末端にオリゴヌクレオチドを結合させる、工程と、

(c) 断片を T4 RNA リガーゼに接触させる工程であって、それによって前記非コード RNA を前記オリゴヌクレオチドにライゲーションし、前記断片化した DNA 分子のクロスリンクを脱クロスリンクする、工程と、

(d) 二本鎖 DNA 断片を作製するために、ライゲーションされた RNA を逆転写酵素で伸長させる工程と、

(e) 前記検出可能な標識に結合する薬剤に結合されたエンドヌクレアーゼに前記二本鎖 DNA 断片を接触させる工程であって、それによって前記検出可能な標識付近の DNA を消化する、工程と、

(f) 配列決定アダプターを (e) の生成物に結合させる工程と、

t を含む、方法。

10

20

30

40

50

【請求項 1 2 0】

前記非コードRNAは長い非コードRNAである、請求項 1 1 9 に記載の方法。

【請求項 1 2 1】

前記オリゴヌクレオチドはアデニル化されている、請求項 1 1 9 に記載の方法。

【請求項 1 2 2】

前記検出可能な標識はビオチンを含む、請求項 1 1 9 に記載の方法。

【請求項 1 2 3】

前記薬剤はストレプトアビジンを含む、請求項 1 1 9 に記載の方法。

【請求項 1 2 4】

前記エンドヌクレアーゼは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む、請求項 1 1 9 に記載の方法。 10

【請求項 1 2 5】

(f) の前記生成物を濃縮する工程をさらに含む、請求項 1 1 9 に記載の方法。

【請求項 1 2 6】

前記二本鎖DNA断片の配列を得る工程をさらに含む、請求項 1 1 9 に記載の方法。

【請求項 1 2 7】

(b) の前に、前記安定化された生体試料をRNase Hに接触させる工程をさらに含む、請求項 1 1 9 に記載の方法。

【請求項 1 2 8】

前記安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である、請求項 1 1 9 ～ 1 2 7 のいずれか 1 つに記載の方法。 20

【請求項 1 2 9】

前記安定化された生体試料は安定化された細胞溶解物を含む、請求項 1 1 9 ～ 1 2 7 のいずれか 1 つに記載の方法。

【請求項 1 3 0】

前記安定化された生体試料は安定化された無傷細胞を含む、請求項 1 1 9 ～ 1 2 7 のいずれか 1 つに記載の方法。

【請求項 1 3 1】

前記安定化された生体試料は安定化された無傷の核を含む、請求項 1 1 9 ～ 1 2 7 のいずれか 1 つに記載の方法。 30

【請求項 1 3 2】

前記オリゴヌクレオチドはバーコードをさらに含む、請求項 1 1 9 ～ 1 3 1 のいずれか 1 つに記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

相互参照

本出願は、2020年12月30日に出願された米国仮特許出願第63/132,304号、2021年2月4日に出願された米国仮特許出願第63/145,859号、2021年6月29日に出願された米国仮特許出願第63/216,170号、2021年6月30日に出願された米国仮特許出願第63/217,009号、2021年7月22日に出願された米国仮特許出願第63/224,722号、および2021年11月19日に出願された米国仮特許出願第63/281,499号の利点を主張するものであり、上記文献の各々は、参照によりその全体が本明細書に組み込まれる。 40

【背景技術】

【0002】

高品質の連続したゲノム配列を得ることは、特に、限定された供給源材料が配列分析に利用可能である場合、しばしば困難である。生配列データを得ることは、より速く、より低コストで利用可能になってきたが、データを効率的かつ正確に分析して組み立てるための好適な方法は、依然として課題である。 50

【発明の概要】

【0003】

一態様では、核酸処理の方法が提供される。場合によっては、方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程と、核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントに切断する工程であって、切断がトランスポザーゼによって達成される、工程と、第1のセグメントおよび第2のセグメントに第1のリコンビナーゼ部位を結合させる工程と、リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、第1のセグメントおよび第2のセグメントを接触させる工程であって、それによって、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含む結合された核酸を生成する、工程とを含む。いくつかの実施形態では、切断する工程および結合させる工程は、協奏反応で生じる。他の実施形態では、切断する工程および結合させる工程は連続して生じる。

10

【0004】

場合によっては、トランスポザーゼはTn5トランスポザーゼである。場合によっては、第2のリコンビナーゼ部位は、直接反復配列(direct repeats)として配向された2つのリコンビナーゼ部位を含む。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、インテグラーゼは、PhiC31インテグラーゼ、Bxb1インテグラーゼ、またはそれらの組み合わせである。

【0005】

場合によっては、方法は、結合された核酸の5'末端を、結合された核酸の3'末端にライゲーションすることによって、結合された核酸を環状化する工程をさらに含む。

20

【0006】

特定の場合において、方法は、結合された核酸の少なくとも一部を配列決定する工程をさらに含む。特定の場合において、配列決定する工程は、第1の配列の少なくとも一部および第2の配列の少なくとも一部を配列決定することを含む。特定の場合において、方法は、第1の配列の少なくとも一部および第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む。特定の場合において、方法は、配列決定からの情報を使用して三次元ゲノム解析を行う工程をさらに含む。

【0007】

様々な場合において、安定化された試料はクロスリンクされた試料である。様々な場合において、安定化された試料を得る工程は、試料を得ることと、試料を安定化することとを含む。様々な場合において、安定化された試料を得る工程は、前もって安定化された試料を得ることを含む。様々な場合において、核酸結合タンパク質は、クロマチンまたはその成分を含む。様々な場合において、安定化された試料は、約62℃にて約0.3% SDSで細胞を処理することによって調製された核を含む。様々な場合において、第1のリコンビナーゼ部位および第2のリコンビナーゼ部位は、attPおよびattBインテグラーゼ部位を含む。様々な場合において、第1のリコンビナーゼ部位は、第2のリコンビナーゼ部位とは異なる。様々な場合において、第1のリコンビナーゼ部位は、attPまたはattBインテグラーゼ部位である。様々な場合において、第2のリコンビナーゼ部位は、attPまたはattBインテグラーゼ部位である。様々な場合において、第1のリコンビナーゼ部位はattPインテグラーゼ部位であり、第2のリコンビナーゼ部位はattBインテグラーゼ部位である。様々な場合において、第1のリコンビナーゼ部位はattBインテグラーゼ部位であり、第2のリコンビナーゼ部位はattPインテグラーゼ部位である。様々な場合において、結合された核酸はヘアピンループを形成しない。様々な場合において、第1のリコンビナーゼ部位および第2のリコンビナーゼ部位は、トランスポザーゼモザイク末端を含む。様々な場合において、リンカー配列はバーコード配列を含む。様々な場合において、バーコード配列は、パーティション起源を示す。様々な場合において、バーコード配列は、細胞起源を示す。様々な場合において、バーコード配列は、細胞集団起源を示す。様々な場合において、バーコード配列は、生物起源を示す。様々

30

40

50

な場合において、バーコード配列は、種起源を示す。様々な場合において、パーティションは、アレイまたはビーズ上のウェル、液滴、または表面位置である。様々な場合において、方法は4時間以下で完了する。様々な場合において、安定化された試料は、50,000個以下の細胞を含む。様々な場合において、安定化された試料は、少なくとも10,000個の細胞を含む。様々な場合において、安定化された試料は安定化された核を含む。様々な場合において、安定化された試料は、50,000個以下の核を含む。様々な場合において、安定化された試料は、少なくとも10,000個の核を含む。様々な場合において、安定化された試料は超音波処理されない。

【0008】

別の態様では、核酸処理の方法が提供され、方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程と、核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントへ切断する工程と、第1のセグメントおよび第2のセグメントに第1のリコンビナーゼ部位を結合させる工程と、リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、第1のセグメントおよび第2のセグメントを接触させる工程であって、それによって、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含む結合された核酸を生成し、第2のリコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む、工程とを含む。いくつかの実施形態では、切断する工程および結合させる工程は、協奏反応で生じる。他の実施形態では、切断する工程および結合させる工程は連続して生じる。別の態様では、核酸処理の方法が提供され、方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程と、核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントへ切断する工程と、第1のセグメントおよび第2のセグメントに第1のリコンビナーゼ部位を結合させる工程と、リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、第1のセグメントおよび第2のセグメントを接触させる工程であって、それによって、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含む結合された核酸を生成する、工程とを含む。いくつかの実施形態では、第2のリコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む。

【0009】

場合によっては、切断する工程はトランスポザーゼによって達成される。場合によっては、トランスポザーゼは、Tn5トランスポザーゼである。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、インテグラーゼは、PhiC31インテグラーゼ、Bxb1インテグラーゼ、またはそれらの組み合わせである。

【0010】

場合によっては、方法は、結合された核酸の5'末端を、結合された核酸の3'末端にライゲーションすることによって、結合された核酸を環状化する工程をさらに含む。

【0011】

特定の場合において、方法は、結合された核酸の少なくとも一部を配列決定する工程をさらに含む。特定の場合において、配列決定する工程は、第1の配列の少なくとも一部および第2の配列の少なくとも一部を配列決定することを含む。特定の場合において、方法は、第1の配列の少なくとも一部および第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む。特定の場合において、方法は、配列決定からの情報を使用して三次元ゲノム解析を行う工程をさらに含む。

【0012】

特定の場合において、安定化された試料はクロスリンクされた試料である。特定の場合において、安定化された試料を得る工程は、試料を得ることと、試料を安定化することとを含む。特定の場合において、安定化された試料を得る工程は、前もって安定化された試料を得ることを含む。特定の場合において、核酸結合タンパク質は、クロマチンまたはその成分を含む。特定の場合において、第1のリコンビナーゼ部位および第2のリコンビナ

ーゼ部位は、a t t P および a t t B インテグラーゼ部位を含む。特定の場合において、第1のリコンビナーゼ部位は、第2のリコンビナーゼ部位とは異なる。特定の場合において、第1のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位である。特定の場合において、第2のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位である。様々な場合において、第1のリコンビナーゼ部位は a t t P インテグラーゼ部位であり、第2のリコンビナーゼ部位は a t t B インテグラーゼ部位である。様々な場合において、第1のリコンビナーゼ部位は a t t B インテグラーゼ部位であり、第2のリコンビナーゼ部位は a t t P インテグラーゼ部位である。特定の場合において、結合された核酸はヘアピンループを形成しない。特定の場合において、第1のリコンビナーゼ部位および第2のリコンビナーゼ部位は、トランスポザーゼモザイク末端を含む。特定の場合において、リンカー配列はバーコード配列を含む。

10

【0013】

様々な場合において、バーコード配列は、パーティション起源を示す。様々な場合において、バーコード配列は、細胞起源を示す。様々な場合において、バーコード配列は、細胞集団起源を示す。様々な場合において、バーコード配列は、生物起源を示す。様々な場合において、バーコード配列は、種起源を示す。様々な場合において、パーティションは、アレイまたはビーズ上のウェル、液滴、または表面位置である。様々な場合において、方法は4時間以下で完了する。様々な場合において、安定化された試料は、50,000個以下の細胞を含む。様々な場合において、安定化された試料は、少なくとも10,000個の細胞を含む。様々な場合において、安定化された試料は安定化された核を含む。様々な場合において、安定化された試料は、約62℃にて約0.3% SDSで細胞を処理することによって調製された核を含む。様々な場合において、安定化された試料は、50,000個以下の核を含む。様々な場合において、安定化された試料は、少なくとも10,000個の核を含む。様々な場合において、安定化された試料は超音波処理されない。

20

【0014】

他の態様では、(a) 少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、(b) 複合体を形成するために核酸分子をデンドリマーと接触させる工程であって、デンドリマーの1つ以上のポリマーが末端の一級アミンを含む、工程と、(c) 核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントへ切断する工程と、(d) 複数のセグメントの第1のセグメントおよび第2のセグメントを接合部で結合させる工程と、を含む方法が提供される。場合によっては、デンドリマーはクロスリンカー (cross linker) で修飾される。場合によっては、方法は、(b) の前に、デンドリマーをクロスリンカーに接触させる工程をさらに含む。場合によっては、クロスリンキング剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、クロスリンカーはソラレン、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファミド (isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン (ciccy cloplatin)、エプタプラチン (eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (mitozolomide)、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンを含む。場合によっては、ソラレンはN-ヒドロキシスクシンイミド (NHS) エステル-コンジュゲートソラレンを含む。場合によっては、デンドリマーはポリアミドアミン (PAMAM) デンドリマーを含む。場合によっては、上記方法はさらに、(e) デンドリマーからクロスリンカーを分離する工程をさらに含む。場合によっては、分離する工程は熱アルカリ処理を含む。場合によっては、分離する工程はUV放射への曝露を含む。場合によっては、複数のセグメントの一部は結合してコンカテマーを形成

30

40

50

する。場合によっては、コンカテマーは少なくとも3個のセグメントを含む。場合によっては、コンカテマーは少なくとも4個のセグメントを含む。場合によっては、コンカテマーは少なくとも5個のセグメントを含む。場合によっては、コンカテマーは少なくとも6個のセグメントを含む。場合によっては、コンカテマーは少なくとも8個のセグメントを含む。場合によっては、コンカテマーは少なくとも10個のセグメントを含む。場合によっては、デンドリマーの分子量は、5キロダルトン(kDa)~125kDaである。場合によっては、デンドリマーの分子量は、6kDa~8kDaである。場合によっては、デンドリマーの分子量は、25kDa~35kDaである。場合によっては、デンドリマーの分子量は、110kDa~125kDaである。場合によっては、デンドリマーは32~512個の反応基を含む。場合によっては、デンドリマーは約32個の反応基を含む。場合によっては、デンドリマーは約128個の反応基を含む。場合によっては、デンドリマーは約512個の反応基を含む。場合によっては、方法は、(b)の後に、デンドリマー複合体を光活性化する工程をさらに含む。場合によっては、方法は、(f)複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程をさらに含む。場合によっては、切断する工程は、核酸分子をデオキシリボヌクレアーゼ(DNase)と接触させることを含む。場合によっては、DNaseは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む。場合によっては、安定化された生体試料はクロスリンク剤で処理される。場合によっては、クロスリンク剤は化学的固定剤である。場合によっては、化学的固定剤は、ホルムアルデヒド、ソラレン、ジスクシンイミジルグルタレート(DSG)、エチレングリコールビス(スクシンイミジルスクシネート)(EGS)、紫外線、またはそれらの組み合わせを含む。場合によっては、クロスリンク剤は、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファミド(isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン(ciccy cloplatin)、エプタプラチン(eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド(mitozolomide)、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキシソルピシン、ダウノルピシン、エピルピシン、またはイダルピシンを含む。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。場合によっては、安定化された生体試料は安定化された無傷細胞を含む。場合によっては、安定化された生体試料は安定化された無傷の核を含む。様々な場合において、安定化された試料は、約62℃にて約0.3%SDSで細胞を処理することによって調製された核を含む。場合によっては、工程(c)は、無傷細胞または無傷の核の溶解の前に実施される。場合によっては、方法は、工程(d)の前に、安定化された生体試料中で細胞および／または核を溶解させる工程をさらに含む。場合によっては、安定化された生体試料は3,000,000未満の細胞を含む。場合によっては、安定化された生体試料は1,000,000未満の細胞を含む。場合によっては、安定化された生体試料は100,000未満の細胞を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端をfish-inすること、および平滑末端をライゲーションすることを含む。場合によっては、結合させる工程は、少なくとも1つの架橋オリゴヌクレオチドに、少なくとも第1のセグメントおよび第2のセグメントを接触させることを含む。場合によっては、架橋オリゴヌクレオチドはバーコード配列を含む。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを、複数の架橋オリゴヌクレオチドに連続して接触させることを含む。場合によっては、結合させる工程は、架橋オリゴヌクレオチドの固有の配列を受け取る安定化された生体試料

10

20

30

40

50

の試料、細胞、核、染色体、または核酸分子をもたらす。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含む。場合によっては、結合させる工程は、パーティションで行われる。場合によっては、パーティションは、アレイまたはビーズ上のウェル、液滴、または表面位置である。場合によっては、上記方法はさらに、(g) 第1のリード対を生成するために接合部の各側で少なくともいくつかの配列を得る工程を含む。場合によっては、上記方法はさらに、(h) 第1のリード対をコンティグのセットにマッピングする工程と、(i) ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とを含む。場合によっては、上記方法はさらに、(h) 第1のリード対をコンティグのセットにマッピングする工程と、(i) コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含む。場合によっては、上記方法はさらに、(h) 第1のリード対をコンティグのセットにマッピングする工程と、(i) コンティグのセット内のバリエーションをフェーズに割り当てる工程とを含む。場合によっては、上記方法はさらに、(h) 第1のリード対をコンティグのセットにマッピングする工程と、(i) コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(j) (1) 安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2) バリエーションの存在に基づいて薬物を選択すること、または、(3) 安定化された生体試料についての薬物効力を同定すること、のうちの1つ以上から選択される工程を行う工程とを含む。

10

20

30

40

【0015】

別の態様では、(a) 少なくとも1つの核酸結合タンパク質に複合体化された核酸分子および少なくとも1つの非コードRNAを含む安定化された生体試料を得る工程と、(b) Tn5トランスポザゼ、ならびに、モザイク末端および検出可能な標識を含むオリゴヌクレオチドに、DNA分子を接触させる工程であって、それによってDNA分子を断片化し、断片化したDNA分子の末端にオリゴヌクレオチドを結合させる、工程と、(c) 断片をT4 RNAリガーゼに接触させる工程であって、それによって非コードRNAをオリゴヌクレオチドにライゲーションし、クロスリンクを脱クロスリンクする(reversing)、工程と、(d) 二本鎖DNA断片を作製するために、ライゲーションされたRNAを逆転写酵素で伸長させる工程と、(e) 検出可能な標識に結合する薬剤に結合されたエンドヌクレアーゼに二本鎖DNA断片を接触させる工程であって、それによって検出可能な標識付近のDNAを消化する、工程と、(f) 配列決定アダプターを(e)の生成物に結合させる工程と、を含む分析方法が提供される。場合によっては、非コードRNAは長い非コードRNAである。場合によっては、オリゴヌクレオチドはアデニル化されている。場合によっては、検出可能な標識はビオチンを含む。場合によっては、薬剤はストレプトアビジンを含む。場合によっては、エンドヌクレアーゼは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む。場合によっては、方法は、(f)の生成物を濃縮する工程をさらに含む。場合によっては、方法は、二本鎖DNA断片の配列を得る工程をさらに含む。場合によっては、方法は、(b)の前に、安定化された生体試料をRNase Hに接触させる工程をさらに含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。場合によっては、安定化された生体試料は安定化された無傷細胞を含む。場合によっては、安定化された生体試料は安定化された無傷の核を含む。場合によっては、オリゴヌクレオチドはバーコードをさらに含む。

【0016】

引用による組み込み

本明細書で言及されるすべての刊行物、特許、および特許出願は、あたかも個々の刊行物、特許、または特許出願が引用によって組み込まれるよう具体的かつ個別に示されるかのように、同じ程度まで引用により本明細書に組み込まれる。

【図面の簡単な説明】

50

【0017】

本発明の特徴および利点のより良い理解は、本発明の原理が用いられる例示的な実施形態を説明する以下の詳細な説明と、添付図面とを引用することによって得られるであろう。

【図1A】不十分に処理された（図1A）および十分に処理された（図1B）安定化された組織試料を示す。

【図1B】不十分に処理された（図1A）および十分に処理された（図1B）安定化された組織試料を示す。

【図2】定量化および断片サイズ分析を示す。

【図3】本開示の様々な実施形態にかかる例示的なコンピュータシステムの様々な構成要素を示す。

10

【図4】本開示の様々な実施形態に関連して使用することができる例示的なコンピュータシステムのアーキテクチャを示すブロック図である。

【図5】本開示の様々な実施形態に関連して使用することができる例示的なコンピュータネットワークを示す図である。

【図6】本開示の様々な実施形態に関連して使用することができる別の例示的なコンピュータシステムのアーキテクチャを示すブロック図である。

【図7】MNase-Cと比較したDNase-Cについてのリード対分離分布のグラフを示す。

【図8】MNase-Cと比較したDNase-Cについての1番染色体について計算された連鎖距離の累積分布のグラフを示す。

20

【図9】MNase消化の様々な時間および条件についての高占有率CTCF結合部位の周りの相対的なリードカバレッジのグラフを示す。

【図10】MNase消化の様々な時間および条件についてのMNase処理された試料の消化パターン、ならびに各試料中のモノヌクレオソーム：ジヌクレオソームの計算された比を示す。

【図11】University of California, Santa Cruz (UCSC) Genome BrowserからのEncyclopedia of DNA Elements (ENCODE)において報告されたピークと比較したChIP-seqおよびHiChIPの結果を示す。

30

【図12】HiChIP試料についてのCTCF結合部位周辺の相対的なリードカバレッジを示す。

【図13】標的タンパク質に関連するリードのピルアップを示すリードカバレッジのグラフ（図11に示す）および遺伝子アノテーションのグラフにわたって提示されるリード対の接触マップを示す。

【図14】図11と同じENCODEピークに対するMNase HiChIP結果の比較であるが、同じ日およびその後の日の試料反復についての比較を示す。

【図15】本開示の様々な実施形態にかかる、各DNAセグメントを結合するために架橋オリゴヌクレオチドを使用する、改良された近接ライゲーションのための例示的なワークフローを示す。

40

【図16】本開示の様々な実施形態にかかる、分割およびプーリング手法を使用する例示的なワークフローを示す。

【図17】本開示の様々な実施形態にかかる、分割およびプーリング手法を使用する例示的なワークフローを示す。

【図18】本開示の様々な実施形態にかかる、分割およびプーリング手法から得られるバーコードと架橋との組み合わせの例を示す。

【図19A】例示的なインテグラーゼ活性の工程を示す。図19Aでは、phiC31プロファージDNAがストレプトマイセス属（*Streptomyces* sp.）に組み込まれていることが分かる。

【図19B】例示的なインテグラーゼ活性の工程を示す。図19Bでは、インテグラーゼ

50

が、細菌ゲノム中のファージ *a t t B* 部位および *a t t P* 配列に結合し、鎖交換を誘発していることが分かる。

【図 19 C】例示的なインテグラーゼ活性の工程を示す。図 19 C では、組み込みが 3 塩基短い 2 つの新しい配列 *a t t L* および *a t t R* に分解していることが分かる。

【図 19 D】例示的なインテグラーゼ活性の工程を示す。図 19 D では、環状リンカーが組み込みに必要でないことが分かる。

【図 20】組み込み部位の内部露出された末端への送達の例として、*D N a s e* 消化クロマチンへのアダプターライゲーションによる *a t t P* 送達を描いている。

【図 21】33 塩基 *a t t B* セグメントに隣接する配列が、例えば、パーティション識別セグメントを使用して交換され得ることを示す。

【図 22】*a t t B* 線状 *D N A* の組み込みが凝集体内ライゲーションを引き起こすことを示す。

【図 23 A】ライブラリー調製アプローチを描いている。図 23 A では、*a t t B* 含有分子に組み込まれたビオチンが見られるが、*a t t P* アダプターはホスホチオレート化ヌクレオチドを保有している。

【図 23 B】ライブラリー調製アプローチを描いている。図 23 B では、ストレプトアビジンプルダウンがビオチン含有分子のみをプルダウンするために使用されることが分かる。

【図 23 C】ライブラリー調製アプローチを描いている。図 23 C では、*a t t P* 特異的増幅が、組み込まれた分子のみを増幅するために使用されることが分かる。

【図 23 D】ライブラリー調製アプローチを描いている。図 23 D では、代替的にまたは組み合わせて、エキソヌクレアーゼ活性が、組み込まれていない分子および非アダプターライゲーション核酸を除去するために使用されることが分かる。

【図 24】近接ライゲーション事象の単一核タグ付けのための組み込みの使用を描いている。

【図 25】（図の上部から下部へ）断片化（例えば、*D N a s e* を用いるなどの酵素消化による）、末端平滑化（*e n d p o l i s h i n g*）および A テーリング、組換え部位のライゲーション（例えば、A テール適合性 *a t t B* 部位）、リンカーとの組換え（例えば、*a t t P* リンカー）、およびクロスリンクの脱クロスリンク（*c r o s s - l i n k r e v e r s a l*）を含む、リコンビナーゼ（例えば、インテグラーゼ）ベースの近接ライゲーションであって、ゲノムの異なる領域に由来する近接ライゲーションされた核酸をもたらすものの例示的な概略図を示している。

【図 26】（図の上部から下部へ）非組換え *g D N A* にライゲーションされた例示的な *E P* オーバーハング *a t t B* 部位、非組換え *g D N A* にライゲーションされた例示的な *a t t B* 部位、*a t t P* 部位を有する例示的な非組換えビオチンリンカー、*g D N A* 上の *a t t B* 部位に組換えられたリンカーの 1 つの末端、および *g D N A* 上の *a t t B* 部位と組換えられたリンカーの両方の末端を含む、リコンビナーゼベースの近接ライゲーションのための例示的な核酸配列を示す。

【図 27】配列決定アダプター部位（例えば、*P 7* および *P 5* 配列決定 *Y* アダプター）を伴う *a t t B* 部位の例示的な設計および核酸配列を、それ自体（上部）で、および非組換え *g D N A* へのライゲーションされた状態（下部）で示している。

【図 28】他のすべての遊離末端とハイブリダイゼーション適合性である一本鎖オーバーハングを有する多くの遊離末端を有する消化された再構成されたクロマチン凝集体を描いている。

【図 29】単一塩基が *f i l l - i n* された状態の図 28 の消化された再構成されたクロマチン凝集体を描いており、各一本鎖オーバーハングを再アニーリングおよび再ライゲーションに不適合にしている。

【図 30】再構成された凝集体の修飾された遊離末端と適合する切断（*p u n c t u a t i o n*）オリゴヌクレオチドとライゲーションされた図 29 の部分的に *f i l l - i n* された消化された再構成されたクロマチン凝集体を描いている。

10

20

30

40

50

【図 3 1】図 3 0 のライゲーション産物とその後の DNA 結合タンパク質からの放出から生じる切断された DNA 分子を示す。各ゲノムセグメントは、その既知の配列によって同定可能な切断オリゴヌクレオチドによって描写される。ゲノムセグメントはすべて、その出発再構成クロマチン凝集体における入力分子のいくつかの領域を表す。したがって、このセットのリードはハプロタイプフェージングされ (haplotype phased)、アセンブリまたはハプロタイプフェーズの再構築に使用することができる。

【図 3 2】Tn5 を使用して核酸を断片化し、AttB タグを付加し、その後核酸末端の Phic31 結合を行う試料調製方法を描いている。精製されたタグ付き核酸は、増幅に基づくライブラリー調製の前に環状化される。

【図 3 3】図 3 2 の方法のための例示的なワークフローを描いている。

【図 3 4】異なる方法を用いて調製されたライブラリーを使用して得られたカバレッジ均一性の比較を描いている。

【図 3 5】異なる方法を用いて調製されたライブラリー特性の比較を描いている。

【図 3 6】異なる方法を用いて調製されたライブラリーの所定のゲノム間隔における接触数の比較を描いている。

【図 3 7】ライブラリー調製におけるプライマーの取り込みの戦略を描いている。

【図 3 8】異なるライブラリー調製方法を用いて得られたロングレンジ情報の比較を描いている。

【図 3 9】長い非コード RNA (lncRNA) 結合部位を同定する方法を描いている。

【発明を実施するための形態】

【0018】

1つの態様において、ロングレンジおよび構造ゲノムの情報を含むゲノム配列の決定、細胞における核酸の物理的立体構造の決定に関連する、ならびにいくつかの他の方法よりも改善された結果を有する核酸の極端にロングレンジのリード対を生成するための、組成物、システム、および方法が本明細書において提供される。本明細書の方法は、最適な結果を達成するために、限定されないが、クロスリンクした核酸のトランスポザーゼ断片化、インテグラーゼ組換え、および低入力要件を含む技術を利用することができる。2020年6月25日に出願された PCT/US2020/039656 は、その全体が参照により本明細書に組み込まれる。

【0019】

トランスポザーゼ断片化クロマチン

1つの態様において、核酸プロセシングの方法が本明細書で提供される。そのような方法は、少なくとも1つの核酸結合タンパク質と複合体形成した核酸分子を含む安定化された試料を得る工程と、核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントに切断する工程を含むことができ、切断する工程はトランスポザーゼによって達成される。本明細書に記載の方法は、第1のセグメントおよび第2のセグメントに第1のリコンビナーゼ部位を結合させる工程をさらに含むことができる。いくつかの実施形態では、切断する工程および結合させる工程は、協奏反応で生じる。他の実施形態では、切断する工程および結合させる工程は連続して生じる。さらに、本明細書に記載の方法は、リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、第1のセグメントおよび第2のセグメントを接触させる工程であって、それによって、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含む結合された核酸を生成する、工程を含むことができる。場合によっては、安定化された試料は超音波処理されない。

【0020】

本明細書の方法の様々な態様において、安定化された核酸を切断する工程は、トランスポザーゼを使用して行われる。場合によっては、切断する工程は、透過処理された細胞において行われる。場合によっては、切断する工程は、透過処理された核において行われる。場合によっては、トランスポザーゼは、Tn5、Tn3、Tn7、sleeping beauty トランスポザーゼ、またはそれらの組み合わせである。場合によっては、ト

10

20

30

40

50

ランスボザーゼは、T n 5 トランスボザーゼである。

【0021】

本明細書の方法の様々な態様において、リコンビナーゼ部位を含むリンカーは、リコンビナーゼの存在下で切断された核酸に接触させることができ、リコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む。場合によっては、直接反復配列として配向されたリコンビナーゼ部位の存在は、得られる産物が安定したヘアピン構造を形成することを防ぎ得る。場合によっては、結合された核酸は、ヘアピンループを形成しない。場合によっては、得られた産物は、リコンビナーゼ部位が逆反復配列として配向されている産物よりも容易に配列決定される。

【0022】

本明細書の方法の様々な態様では、切断された核酸の第1のセグメントおよび第2のセグメントを、リコンビナーゼの存在下でインテグラーゼ部位を含むリンカーに接触させる。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、インテグラーゼは、P h i C 3 1 インテグラーゼ、B x b 1 インテグラーゼ、またはそれらの組み合わせである。

【0023】

様々な態様において、リコンビナーゼ産物が環状化される、核酸を処理する方法が提供される。場合によっては、リコンビナーゼの末端を除去して、ライゲーションの前に第1のセグメントおよび第2のセグメントを露出させる。いくつかの実施形態では、環状化産物は、配列決定ライブラリーを作製するためにP C Rを使用して増幅される。この方法の例を図32および図33に例示する。

【0024】

本明細書の方法の様々な態様において、リンカーは、モザイク末端、配列決定アダプター、およびa t t B 配列を含む。代替的に、リンカーはモザイク末端および配列決定アダプターを含み、a t t B 配列は、組換えの前に、例えば、リガーゼを用いてトランスボザーゼ産物に付加される。

【0025】

本明細書の方法の特定の態様では、方法は、本明細書で提供される方法などの任意の適切な方法を介して結合された核酸の少なくとも一部を配列決定する工程をさらに含むことができる。場合によっては、配列決定する工程は、第1の配列の少なくとも一部および第2の配列の少なくとも一部を配列決定することを含み得る。特定の場において、方法は、第1の配列の少なくとも一部および第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む得る。様々な場合において、方法は、配列決定からの情報を使用して三次元ゲノム解析を行う工程をさらに含む得る。

【0026】

本明細書の方法の様々な態様において、安定化された試料は、クロスリンクされた試料であり得る。場合によっては、安定化された試料はクロスリンクされた細胞であり得る。場合によっては、安定化された試料はクロスリンクされた核であり得る。場合によっては、安定化された試料はクロスリンクされたクロマチンであり得る。場合によっては、安定化された試料を得る工程は、試料を得ることと、試料を安定化することを含むことができる。場合によっては、安定化された試料を得る工程は、前もって安定化された試料を得ることを含むことができる。場合によっては、核酸結合タンパク質は、クロマチンまたはその成分を含むことができる。

【0027】

本明細書の方法の様々な態様において、レコンビナーゼ部位は、a t t P および a t t B インテグラーゼ部位を含むことができる。場合によっては、第1のリコンビナーゼ部位は、第2のリコンビナーゼ部位とは異なり得る。場合によっては、第1のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位であり得る。場合によっては、第2のリコンビナーゼ部位は、a t t P または a t t B インテグラーゼ部位であり得る。様々な場合において、第1のリコンビナーゼ部位は a t t P インテグラーゼ部位であり、第2

10

20

30

40

50

のリコンビナーゼ部位はa t t Bインテグラーゼ部位である。様々な場合において、第1のリコンビナーゼ部位はa t t Bインテグラーゼ部位であり、第2のリコンビナーゼ部位はa t t Pインテグラーゼ部位である。場合によっては、第1のリコンビナーゼ部位および第2のリコンビナーゼ部位は、トランスポザーゼモザイク末端を含むことができる。

【0028】

本明細書の方法の様々な態様では、リンカーはさらなる配列を含むことができる。場合によっては、リンカー配列はバーコード配列を含むことができる。場合によっては、バーコード配列は、パーティション起源を示すことがある。場合によっては、バーコード配列は、細胞起源を示すことがある。場合によっては、バーコード配列は、細胞集団起源を示すことがある。場合によっては、バーコード配列は、生物起源を示すことがある。場合によっては、バーコード配列は、種起源を示すことがある。場合によっては、リンカーはアダプターを含むことができる。場合によっては、アダプターはP5配列を含むことができる。場合によっては、アダプターはP7配列を含むことができる。

10

【0029】

本明細書の方法の様々な態様において、方法は、1日未満で完了することがある。場合によっては、方法は8時間未満で完了することがある。場合によっては、方法は6時間未満で完了することがある。場合によっては、方法は4時間以下で完了することがある。場合によっては、方法は4～6時間で完了することがある。場合によっては、方法は4～8時間で完了することがある。場合によっては、方法は3～4時間で完了することがある。

【0030】

20

本明細書の方法の様々な態様では、方法は、試料材料の非常に低い入力が必要とし得る。場合によっては、安定化された試料は、50,000個以下の細胞を含むことができる。場合によっては、試料は、40,000個以下の細胞を含むことができる。場合によっては、試料は、30,000個以下の細胞を含むことができる。場合によっては、試料は、20,000個以下の細胞を含むことができる。場合によっては、試料は少なくとも10,000個の細胞を含むことができる。場合によっては、試料は少なくとも20,000個の細胞を含むことができる。場合によっては、試料は少なくとも30,000個の細胞を含むことができる。場合によっては、試料は少なくとも40,000個の細胞を含むことができる。場合によっては、試料は約10,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約20,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約30,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約40,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約10,000個の細胞～約40,000個の細胞を含むことができる。場合によっては、試料は約10,000個の細胞～約30,000個の細胞を含むことができる。場合によっては、試料は約10,000個の細胞～約20,000個の細胞を含むことができる。場合によっては、試料は約20,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約20,000個の細胞～約40,000個の細胞を含むことができる。場合によっては、試料は約20,000個の細胞～約30,000個の細胞を含むことができる。場合によっては、試料は約30,000個の細胞～約50,000個の細胞を含むことができる。場合によっては、試料は約30,000個の細胞～約40,000個の細胞を含むことができる。

30

40

【0031】

本明細書の方法の様々な態様では、安定化された試料は核を含むことがある。場合によっては、安定化された試料は、50,000個以下の核を含むことがある。場合によっては、試料は、40,000個以下の核を含むことがある。場合によっては、試料は、30,000個以下の核を含むことがある。場合によっては、試料は、20,000個以下の核を含むことがある。場合によっては、試料は少なくとも10,000個の核を含むことがある。場合によっては、試料は少なくとも20,000個の核を含むことがある。場合によっては、試料は少なくとも30,000個の核を含むことがある。場合によっては、

50

試料は少なくとも40,000個の核を含むことがある。場合によっては、試料は約10,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約20,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約30,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約40,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約10,000個の核～約40,000個の核を含むことがある。場合によっては、試料は約10,000個の核～約30,000個の核を含むことがある。場合によっては、試料は約10,000個の核～約20,000個の核を含むことがある。場合によっては、試料は約20,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約20,000個の核～約40,000個の核を含むことがある。場合によっては、試料は約20,000個の核～約30,000個の核を含むことがある。場合によっては、試料は約30,000個の核～約50,000個の核を含むことがある。場合によっては、試料は約30,000個の核～約40,000個の核を含むことがある。

10

【0032】

直接反復配列として配向されたリコンビナーゼ部位

別の態様では、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された試料を得る工程を含む核酸処理の方法が提供される。次に、方法は、核酸分子を、少なくとも第1のセグメントおよび第2のセグメントを含む複数のセグメントに切断する工程と、第1のリコンビナーゼ部位を第1のセグメントおよび第2のセグメントに結合させる工程とを含むことができる。それから、方法は、リコンビナーゼの存在下で、第2のリコンビナーゼ部位を含むリンカーに、第1のセグメントおよび第2のセグメントを接触させる工程であって、それによって、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含む結合された核酸を生成し、第2のリコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む、工程を含む。場合によっては、安定化された試料は超音波処理されないことがある。

20

【0033】

本明細書の方法の様々な態様において、安定化された核酸を切断する工程は、トランスポザーゼを使用して行われることがある。場合によっては、切断する工程は、透過処理された細胞において行われることがある。場合によっては、切断する工程は、透過処理された核において行われることがある。場合によっては、トランスポザーゼは、Tn5、Tn3、Tn7、sleeping beautyトランスポザーゼ、またはそれらの組み合わせであり得る。場合によっては、トランスポザーゼは、Tn5トランスポザーゼであり得る。

30

【0034】

本明細書の方法の様々な態様において、リコンビナーゼ部位を含むリンカーは、リコンビナーゼの存在下で切断された核酸に接触させてもよく、リコンビナーゼ部位は、直接反復配列として配向された2つのリコンビナーゼ部位を含む。場合によっては、直接反復配列として配向されたリコンビナーゼ部位の存在は、得られる産物が安定したヘアピン構造を形成することを防ぎ得る。場合によっては、結合された核酸は、ヘアピンループを形成しない。場合によっては、得られた産物は、リコンビナーゼ部位が逆反復配列として配向されている産物よりも容易に配列決定され得る。

40

【0035】

本明細書の方法の様々な態様では、切断された核酸の第1のセグメントおよび第2のセグメントを、リコンビナーゼの存在下でインテグラーゼ部位を含むリンカーに接触させてもよい。場合によっては、リコンビナーゼはインテグラーゼであり得る。場合によっては、インテグラーゼは、PhiC31インテグラーゼ、Bxb1インテグラーゼ、またはそれらの組み合わせであり得る。

【0036】

本明細書の方法の様々な態様では、方法は、本明細書で提供される方法などの任意の適

50

切な方法を介して結合された核酸の少なくとも一部を配列決定する工程をさらに含むことができる。場合によっては、配列決定する工程は、第1の配列の少なくとも一部および第2の配列の少なくとも一部を配列決定することを含むことができる。場合によっては、方法は、第1の配列の少なくとも一部および第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む。場合によっては、方法は、配列決定からの情報を使用して三次元ゲノム解析を行う工程をさらに含む。

【0037】

本明細書の方法の様々な態様において、安定化された試料は、クロスリンクされた試料であり得る。場合によっては、安定化された試料はクロスリンクされた細胞であり得る。場合によっては、安定化された試料はクロスリンクされた核であり得る。場合によっては、安定化された試料はクロスリンクされたクロマチンであり得る。場合によっては、安定化された試料を得る工程は、試料を得ることと、試料を安定化することを含む。場合によっては、安定化された試料を得る工程は、前もって安定化された試料を得ることを含む。場合によっては、核酸結合タンパク質は、クロマチンまたはその成分を含む。

10

【0038】

本明細書の方法の様々な態様において、レコンビナーゼ部位は、attPおよびattBインテグラーゼ部位を含むことができる。場合によっては、第1のレコンビナーゼ部位は、第2のレコンビナーゼ部位とは異なり得る。場合によっては、第1のレコンビナーゼ部位は、attPまたはattBインテグラーゼ部位であり得る。場合によっては、第2のレコンビナーゼ部位は、attPまたはattBインテグラーゼ部位であり得る。様々な場合において、第1のレコンビナーゼ部位はattPインテグラーゼ部位であり、第2のレコンビナーゼ部位はattBインテグラーゼ部位である。様々な場合において、第1のレコンビナーゼ部位はattBインテグラーゼ部位であり、第2のレコンビナーゼ部位はattPインテグラーゼ部位である。場合によっては、第1のレコンビナーゼ部位および第2のレコンビナーゼ部位は、トランスポザーゼモザイク末端を含み得る。

20

【0039】

本明細書の方法の様々な態様では、リンカーはさらなる配列を含み得る。場合によっては、リンカー配列はバーコード配列を含み得る。場合によっては、バーコード配列は、パーティション起源を示すことがある。場合によっては、バーコード配列は、細胞起源を示すことがある。場合によっては、バーコード配列は、細胞集団起源を示すことがある。場合によっては、バーコード配列は、生物起源を示すことがある。場合によっては、バーコード配列は、種起源を示すことがある。場合によっては、リンカーはアダプターを含むことができる。場合によっては、アダプターはP5配列を含むことができる。場合によっては、アダプターはP7配列を含むことができる。

30

【0040】

本明細書の方法の様々な態様において、方法は、1日未満で完了することがある。場合によっては、方法は8時間未満で完了することがある。場合によっては、方法は6時間未満で完了することがある。場合によっては、方法は4時間以下で完了することがある。場合によっては、方法は4～6時間で完了することがある。場合によっては、方法は4～8時間で完了することがある。場合によっては、方法は3～4時間で完了することができる。

40

【0041】

本明細書の方法の様々な態様では、方法は、試料材料の非常に低い入力が必要とすることができる。場合によっては、安定化された試料は、50,000個以下の細胞を含むことがある。場合によっては、試料は、40,000個以下の細胞を含むことがある。場合によっては、試料は、30,000個以下の細胞を含むことがある。場合によっては、試料は、20,000個以下の細胞を含むことがある。場合によっては、試料は少なくとも10,000個の細胞を含むことができる。場合によっては、試料は少なくとも20,000個の細胞を含むことができる。場合によっては、試料は少なくとも30,000個の細胞を含むことができる。場合によっては、試料は少なくとも約40,000個の細胞を含むことができる。

50

含むことができる。場合によっては、試料は約 10,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 20,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 30,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 40,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 10,000 個の細胞～約 40,000 個の細胞を含むことができる。場合によっては、試料は約 10,000 個の細胞～約 30,000 個の細胞を含むことができる。場合によっては、試料は約 10,000 個の細胞～約 20,000 個の細胞を含むことができる。場合によっては、試料は約 20,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 20,000 個の細胞～約 40,000 個の細胞を含むことができる。場合によっては、試料は約 20,000 個の細胞～約 30,000 個の細胞を含むことができる。場合によっては、試料は約 30,000 個の細胞～約 50,000 個の細胞を含むことができる。場合によっては、試料は約 30,000 個の細胞～約 40,000 個の細胞を含むことができる。

10

【0042】

本明細書の方法の様々な態様では、安定化された試料は核を含むことがある。場合によっては、安定化された試料は、50,000 個以下の核を含むことがある。場合によっては、試料は、40,000 個以下の核を含むことがある。場合によっては、試料は、30,000 個以下の核を含むことがある。場合によっては、試料は、20,000 個以下の核を含むことがある。場合によっては、試料は少なくとも 10,000 個の核を含むことがある。場合によっては、試料は少なくとも 20,000 個の核を含むことがある。場合によっては、試料は少なくとも 30,000 個の核を含むことがある。場合によっては、試料は少なくとも 40,000 個の核を含むことがある。場合によっては、試料は約 10,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 20,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 30,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 40,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 10,000 個の核～約 40,000 個の核を含むことができる。場合によっては、試料は約 10,000 個の核～約 30,000 個の核を含むことができる。場合によっては、試料は約 10,000 個の核～約 20,000 個の核を含む。場合によっては、試料は約 20,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 20,000 個の核～約 40,000 個の核を含むことができる。場合によっては、試料は約 20,000 個の核～約 30,000 個の核を含むことができる。場合によっては、試料は約 30,000 個の核～約 50,000 個の核を含むことができる。場合によっては、試料は約 30,000 個の核～約 40,000 個の核を含むことができる。

20

30

【0043】

コンカテマーを作製するための近接ライゲーション

本明細書では、近接ライゲーションを使用してコンカテマー形成を可能にする組成物、システム、および方法が提供される。例えば、核酸結合タンパク質と複合体化された核酸分子を有する安定化された生体試料などの生体試料を、デンドリマーと接触させて複合体を形成することができる。別の例では、生体試料は、デンドリマーと接触させることによって安定化させて複合体を形成することができる。次に、核酸分子を、複数のセグメント、例えば、少なくとも第 1 のセグメントおよび第 2 のセグメントに切断することができる。その後、複数のセグメントを複数の接合部に結合させることができ、例えば、第 1 のセグメントおよび第 2 のセグメントを接合部に結合させることができる。

40

【0044】

本明細書の方法の特定の態様では、生体試料、例えば、核酸結合タンパク質およびデンドリマーと複合体化された核酸分子を有する安定化された生体試料。場合によっては、デンドリマーはソラレンとコンジュゲートされる。場合によっては、デンドリマーはアジド

50

ー P e g 4 - N - ヒドロキスクシンイミド (N H S) エステルとコンジュゲートされる。場合によっては、アジドー P e g 4 - N H S エステルの N H S エステルは、デンドリマー上の一級アミンと反応して、反応性アジド基を有するデンドリマーを生じさせる。場合によっては、カルボキシル化ビーズ (例えば、磁気ビーズ) は、1 - エチル - 3 - (3 - ジメチルアミノプロピル) カルボジイミド (E D C) / スルホ - N H C 化学を用いて、ジベンゾシクロオクチン - アミン (D B C O) - P e g 4 - アミンビルディングブロックとコンジュゲートすることによって調製される。これらの調製されたビーズは、例えば、近接ライゲーションの前に磁気分離法によってデンドリマーを単離するために使用することができる。

【 0 0 4 5 】

場合によっては、デンドリマーは化合物で修飾されるか、または化合物と接触させる。例えば、場合によっては、デンドリマーはソラレンで修飾される。場合によっては、ソラレンは N - ヒドロキスクシンイミド (N H S) エステル - コンジュゲートソラレンを含む。場合によっては、デンドリマーはポリアミドアミン (P A M A M) デンドリマーを含む。場合によっては、デンドリマーは、クロスリンク剤、例えば、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス (2 - クロロエチル) エチルアミン、ビス (2 - クロロエチル) メチルアミン、トリス (2 - クロロエチル) アミン、イソファミド (i s o f a m i d e) 、カルムスチン、ロムスチン、ストレプトゾシン、プスルファン、シスプラチン、カルボプラチン、シシクロプラチン (c i c y c l o p l a t i n) 、エプタプラチン (e p t a p l a t i n) 、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (m i t o z o l o m i d e) 、テモゾロミド、マイトマイシン C 、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンで修飾される。場合によっては、デンドリマーは、挿入剤、抗生物質、または副溝結合剤で修飾される。

【 0 0 4 6 】

本明細書の方法は、デンドリマーから化合物を分離する工程を含むことができる。例えば、ソラレンなどの化合物は、熱を使用してデンドリマーから分離することができる。場合によっては、ソラレンなどの化合物は、アルカリ条件または高 p H を使用して、デンドリマーから分離される。代替的に、ソラレンなどの化合物は、熱およびアルカリ条件を使用して、デンドリマーから分離される。化合物 (例えば、ソラレン) はさらに、U V 放射を使用してデンドリマーから分離され得る。

【 0 0 4 7 】

任意の適切なデンドリマーを、本明細書の方法において使用することができる。デンドリマーの分子量は、約 5 キロダルトン (k D a) ~ 約 1 2 5 k D a であり得る。場合によっては、デンドリマーの分子量は、6 k D a ~ 8 k D a である。場合によっては、デンドリマーの分子量は、2 5 k D a ~ 3 5 k D a である。場合によっては、デンドリマーの分子量は、1 1 0 k D a ~ 1 2 5 k D a である。場合によっては、デンドリマーは 3 2 ~ 5 1 2 個の反応基を含む。場合によっては、デンドリマーは約 3 2 個の反応基を含む。場合によっては、デンドリマーは約 1 2 8 個の反応基を含む。場合によっては、デンドリマーは約 5 1 2 個の反応基を含む。場合によっては、デンドリマーは G e n 3 デンドリマーである。場合によっては、デンドリマーは G e n 5 デンドリマーである。場合によっては、デンドリマーは G e n 7 デンドリマーである。

【 0 0 4 8 】

本明細書の方法は、セグメントの少なくとも一部をコンカテマーに結合させることができる。例えば、コンカテマーを形成するために、少なくとも 2 個のセグメント、少なくとも 3 個のセグメント、少なくとも 4 個のセグメント、少なくとも 5 個のセグメント、少なくとも 6 個のセグメント、少なくとも 7 個のセグメント、少なくとも 8 個のセグメント、少なくとも 9 個のセグメント、少なくとも 1 0 個のセグメント、またはそれ以上を結合さ

ることができる。場合によっては、オリゴヌクレオチドは、各セグメント間に結合される。場合によっては、オリゴヌクレオチドは架橋オリゴヌクレオチドである。場合によっては、オリゴヌクレオチドはアダプターオリゴヌクレオチドである。場合によっては、オリゴヌクレオチドは切断オリゴヌクレオチドである。場合によっては、架橋オリゴヌクレオチド、アダプターオリゴヌクレオチド、および／または切断オリゴヌクレオチドは、バーコード配列を含む。場合によっては、架橋オリゴヌクレオチド、アダプターオリゴヌクレオチド、および／または切断オリゴヌクレオチドは、ジベンゾーシクロオクチン（D B C O）部分で修飾される。場合によっては、D B C O部分は、銅フリークリックケミストリー（copper free click chemistry）を促進する。場合によっては、複数のオリゴヌクレオチドは、各セグメント間に連続して結合される。結合は、オリゴヌクレオチド（例えば、架橋オリゴヌクレオチド）の固有の配列を受け取る安定化された生体試料の試料、細胞、核、染色体、または核酸分子をもたらすことができる。

10

【0049】

場合によっては、デンドリマーを安定化された生体試料と接触させて複合体を形成した後、例えば、約360nmの波長を有するUV放射線に複合体を曝露することによって複合体を光活性化し、それによってクロスリンクした複合体を生成する。場合によっては、クロスリンクは、核酸上に付加物を残すことなく脱クロスリンク可能（reversible）である。

【0050】

本明細書の方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程をさらに含むことができる。本明細書の方法は、セグメントサイズの任意の適切な範囲を含むことができる。。

20

【0051】

本明細書に提供される方法における切断は、任意の適切な方法を使用して、例えば、ヌクレアーゼまたはデオキシリボヌクレアーゼ（DNase）を使用することによって行うことができる。場合によっては、DNaseは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む。

【0052】

本明細書の方法の安定化された生体試料は、安定化剤またはクロスリンク試薬を用いて処理されることによって安定化させることができる。場合によっては、クロスリンク剤は、化学的固定剤、例えば、ホルムアルデヒド、ソラレン、ジスクシンイミジルグルタレート（DSG）、エチレングリコールビス（スクシンイミジルスクシネート）（EGS）、紫外線、またはそれらの組み合わせである。場合によっては、クロスリンク剤は、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス（2-クロロエチル）エチルアミン、ビス（2-クロロエチル）メチルアミン、トリス（2-クロロエチル）アミン、イソファミド（isofamide）、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン（cycloplatin）、エプタプラチン（eptaplatin）、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド（mitozolomide）、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキシソルピシン、ダウノルビシン、エピルビシン、またはイダルビシンを含む。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料であり得る。場合によっては、安定化された生体試料は安定化された無傷細胞または安定化された無傷の核を含む。場合によっては、方法は、安定化された生体試料中の細胞および／または核を溶解する工程を含む。本明細書の方法の切断する工程は、無傷細胞または無傷の核の溶解の前に行うことができる。

30

40

【0053】

本明細書の方法は、少数の細胞を含む、安定化された生体試料上で行うことができる。

50

例えば、場合によっては、安定化された生体試料は3,000,000個未満の細胞を含む。安定化された生体試料は、約1,000,000個未満の細胞、約500,000個未満の細胞、約400,000個未満の細胞、約300,000個未満の細胞、約200,000個未満の細胞、約100,000個未満の細胞、またはそれ未満を含むことができる。

【0054】

本明細書の方法の態様では、方法は、第1のリード対を生成するために、接合部の各側で少なくともいくつかの配列を得る工程をさらに含むことができる。加えて、本方法は、第1のリード対をコンティグのセットにマッピングする工程と、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とをさらに含むことができる。代替的に、または組み合わせて、本方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含むことができる。代替的に、または組み合わせて、本方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットにおけるバリエーションをフェーズに割り当てる工程とを含む。代替的に、または組み合わせて、本方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、バリエーションの存在に基づいて薬物を選択すること、または、安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を行う工程とをさらに含むことができる。

【0055】

本明細書の方法の態様では、近接ライゲーションは、コンカテマーの各セグメント間に結合したDBC O修飾架橋オリゴヌクレオチドを用いるなどの、銅フリークリックケミストリーを含む、クリックケミストリーを用いて行うことができる。その後、コンカテマーを、例えば、デンドリマーを介して結合することができる。ライゲーションした分子を濃縮するために、架橋オリゴヌクレオチドの特徴を標的とすることができる。ある例では、DBC O含有オリゴヌクレオチドを、ビーズなどのストレプトアビジン基質で単離することができるアジド-ビオチン部分と反応させることができる。他の例では、DBC O含有オリゴヌクレオチドを、ジスルフィド結合を含むアジド修飾NHSSSS-dPEG4-ビオチンと反応させることができ、アジド-PEG3アミンを使用して、アジドをNHSSSS-dPEG4-ビオチンに添加することができ、ライブラリー調製のために核酸を単離するために、このジスルフィド結合は、DTTと加熱を使用して、例えば、約10分間70℃で加熱することによって、還元することができる。

【0056】

本明細書の方法の態様では、核酸断片が接触しているデンドリマーは、核酸断片の近接ライゲーションの前に、試料中の残りの核酸から分離または単離させることができる。この工程は、近接ライゲーションによって形成されたコンカテマーが、同じデンドリマーと接触した断片を確実に含むようにすることができる。このことは、所定のコンカテマーのすべてのセグメントが、元の安定化された試料中で互いに近接していたことを意味することができる。したがって、どの核酸領域がどの他の領域に近接していたかについての単なる対情報ではなく、そのようなアプローチは、はるかに複雑な近接情報、例えば、3、4、5、6、7、8、9、10、またはそれ以上の核酸領域がすべて互いに近接していたという情報をもたらすことができる。

【0057】

場合によっては、核酸断片が接触しているデンドリマーを、残りの核酸から分離または単離させることで、近接ライゲーションの代わりに、これらの断片のバーコード化またはタグ付けを可能にすることができる。所定のデンドリマーと関連付けられる断片は、例えば、液滴またはウェル中でバーコード化またはタグ付けすることができる。配列決定後、配列は、それらのバーコードに基づいて関連付けることができ、近接情報は、上記と同じ

コンカテマー中に存在することではなく、バーコードに基づいて導き出すことができる。この近接情報は、本明細書で論じるように使用することができる。一例では、デンドリマーは、試料中の核酸に複合体化されることで、それらを安定化させ、その後、核酸を断片化し、次いで、デンドリマーをそれらの複合体化された核酸断片と共に単離して、液滴中に封入し、液滴中の核酸は、液滴特異的なバーコードまたは標識で標識され、次いで、核酸を配列決定し、バーコードまたは標識情報を使用して、試料中で互いに近接していた断片を関連付ける。

【0058】

長い非コードRNA分析

長い非コードRNA結合部位を分析する方法が本明細書で提供される。場合によっては、そのような方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子、および少なくとも1つの非コードRNAを含む安定化された生体試料を得る工程を含む。次に、方法は、Tn5トランスポザゼ、ならびに、モザイク末端および検出可能な標識を含むオリゴヌクレオチドに、DNA分子を接触させる工程であって、それによってDNA分子を断片化し、断片化したDNA分子の末端にオリゴヌクレオチドを結合させる、工程を含むことができる。断片をT4 RNAリガーゼに接触させ、それによって非コードRNAをオリゴヌクレオチドにライゲーションし、クロスリンクを脱クロスリンクすることができる。その後、ライゲーションしたRNAを逆転写酵素で伸長させることで、二本鎖DNA断片を作製することができる。次いで、二本鎖DNA断片を、検出可能な標識に結合する薬剤に結合されたエンドヌクレアーゼと接触させ、それによって検出可能な標識付近のDNAを消化することができる。その後、配列決定ライブラリーを作製するために、配列決定アダプターを結合させることができる。場合によっては、非コードRNAへのライゲーションを促進するために、オリゴヌクレオチドは、一方の末端でアデニル化される。場合によっては、オリゴヌクレオチドはバーコードをさらに含む。場合によっては、安定化された生体試料を、トランスポザゼ処理の前に、RNase Hと接触させる。

【0059】

本明細書の方法の態様において、場合によっては、非コードRNAは長い非コードRNAである。場合によっては、非コードRNAはエンハンサーである。場合によっては、非コードRNAはmiRNAである。場合によっては、非コードRNAはY RNAである。場合によっては、非コードRNAはRNase Pである。場合によっては、非コードRNAはpiRNAである。場合によっては、非コードRNAはXistである。

【0060】

本明細書の方法の態様において、場合によっては、検出可能な標識は、クリックケミストリー反応が可能な修飾されたヌクレオチドを含む。場合によっては、検出可能な標識はビオチンを含む。場合によっては、薬剤は抗体、プロテインA、プロテインG、またはストレプトアビジンを含む。場合によっては、非コードRNAに結合されたDNAは、さらなる分析の前に濃縮される。

【0061】

本明細書の方法の態様では、エンドヌクレアーゼは、分析の前に外来性の試料DNAを切断するために使用される。場合によっては、エンドヌクレアーゼは、DNase I、DNase II、小球菌ヌクレアーゼ、制限エンドヌクレアーゼ、またはそれらの組み合わせを含む。

【0062】

本明細書の方法の態様では、配列は、非コードRNAを含有している二本鎖DNA断片で得られる。本明細書に記載される方法をさらに含む任意の適切な配列決定方法を使用することができる。

【0063】

様々な適切な安定化された生体試料は、本明細書の方法で使用するために企図される。本明細書で詳細に別記された安定化された生体試料は、クロスリンク剤、例えば、固定剤、またはUV光を用いてクロスリンクされている。例えば、場合によっては、安定化

された生体試料はクロスリンクしたパラフィン包埋した組織試料である。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。場合によっては、安定化された生体試料は安定化された無傷細胞を含む。場合によっては、安定化された生体試料は安定化された無傷の核を含む。

【0064】

核酸立体構造の評価

第2の細胞または細胞集団の物理的な立体構造と区別可能な細胞、例えば、単一細胞または細胞集団における核酸の物理的な立体構造の決定に関連する組成物、システム、および方法が本明細書に開示される。本明細書の開示の実施を通して、三次元核酸相対位置を示す核酸分子を生成し、任意選択で、タグ（例えば、核酸バーコード）を提供することで、複数の分子に共通の細胞または起源集団を識別することができる。

10

【0065】

本明細書で開示される方法の実施を通して、核酸は、細胞におけるその三次元配置のすべてまたは少なくともいくつかを保存するように得ることができる。このような核酸の露出した核酸ループを切断して、互いにランダムに再結合した内部セグメント末端を露出させ、物理的に近接した露出した末端が互いに結合する（近接結合）可能性がより高くなるようにすることができる。したがって、どの露出された末端が互いに対して結合するようになるかを決定することによって、天然の細胞配置における末端隣接核酸の物理的な近接に関する有用なデータを得ることが可能である。

【0066】

関連するアプローチは、例えば、2016年9月6日にDekkerらに公開されたUS9434985B2で開示されており、この文献は参照により全体として本明細書に組み込まれる。

【0067】

本明細書に開示される方法の実施を通して、ペアエンドのライブラリー成分は、集団の個々の細胞間の立体構造の差異が細胞の集団について容易に識別されるように、または細胞の第1の集団と細胞の第2の集団との間の立体構造の差異が、たとえそれらが同時に分析される場合でも、容易に識別されるように、細胞起源を示す配列情報とともにさらにタグ付けされ得るか、それ以外の方法で提供され得る。タグは、例えば、核酸バーコードを含むことができる。場合によっては、タグは、ゲノム中で隣接していない2つの核酸セグメント間の接合部を含むことができる。核酸分子は、完全または部分的に配列決定される場合に、しばしば各ゲノム末端をそのゲノム遺伝子座にマッピングするのに十分な少なくともいくつかのゲノム配列が得られ、さらに、正確なまたは可能性の高い起源の細胞または細胞集団を同定するのに十分なタグ付け配列または連鎖配列を得られるように、作製することができる。したがって、互いに物理的に近接しているゲノムの2つの領域の有用な配列情報を得ると同時に、この物理的な立体構造が生じる細胞または細胞集団の有用な情報も得て、それを、その細胞または細胞集団において同時に生じる他の物理的な立体構造情報の文脈で評価することができるようになる。

20

30

【0068】

細胞中のゲノム核酸または他の核酸を安定化させることができ、真核生物細胞については、本明細書に組み込まれるかまたはそうでなければ公知である方法などの公知の方法に従って、核を任意選択で単離させる。例えば、図1Aおよび図1Bでは、処理された安定化された組織試料が例示されている。図1Aは、不十分に処理された組織試料を例示している。図1Bは、十分に処理された安定化された組織試料を例示している。

40

【0069】

本明細書の開示と一致する核酸は、任意の数の細胞核酸、例えば、原核生物の初代ゲノムもしくはプラスミド核酸、真核生物の核、ミトコンドリアもしくはプラスチド核酸、または、場合によっては、試料中の細胞質核酸、例えば、rRNA、mRNA、もしくは外因性核酸、例えば、試料のウイルスもしくは他の病原体もしくは他の外因性核酸を含む。

【0070】

50

安定化された核酸は、場合によっては、少なくともいくつかの核酸が個々のパーティションに分布するように分布され得る。例示的なパーティションとしては、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【0071】

安定化された核酸は、特定の細胞についての核酸配置情報を得るために、後の再結合のために内部切断部を露出するように断片化することができる。多くの断片化アプローチが公知であり、本明細書の開示と一致している。核酸は、制限エンドヌクレアーゼ、プログラム可能なエンドヌクレアーゼ、例えば、ガイドRNAに結合したCRISPR/Cas分子、非特異的エンドヌクレアーゼ（例えば、DNase）、タグメンテーション（tagmentation）、せん断、超音波処理、加熱、または他の機構の1つまたは複数の集団を使用して断片化することができる。場合によっては、DNaseは非配列特異的である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性である。場合によっては、DNaseは二本鎖DNAに特異的である。場合によっては、DNaseは二本鎖DNAよりも優先される。場合によっては、DNaseは一本鎖DNAに特異的である。場合によっては、DNaseは、一本鎖DNAよりも優先される。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseは小球菌ヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼのうちの1つ以上から選択される。他の適切なヌクレアーゼも本開示の範囲内である。

【0072】

特に、2014年8月7日にGreenらに公開されたWO2014121091A1（後に2015年12月17日にUS20150363550A1として公開され、2018年10月2日にUS10089437B2として公開される）の開示は、全体として本明細書に組み込まれる。同様に、2016年2月4日にFieldsらに公開されたWO2016019360A1（後に2017年11月23日にUS20170335369A1として公開された）の開示は、全体として本明細書に組み込まれる。同様に、2017年8月31日にGreenらに公開されたWO2017147279A1の開示は、全体として本明細書に組み込まれる。

【0073】

核酸は、結合の前または後に表面に結合することができる。例示的な表面としては、限定されないが、ビーズ、アレイ、およびウェルが挙げられる。場合によっては、表面は、SPRIビーズなどの固相可逆的固定化（SPRI）表面である。結合前に核酸を表面に結合させることにより、下流工程の性能を改善することができ、例えば、染色体間ライゲーションまたは結合を減少させ、染色体内ライゲーションまたは結合を増加させることができる。

【0074】

核酸は、結合の前または後に免疫沈降させてもよい。そのような方法は、クロマチンを断片化する工程、および、断片を、アセチル化ヒストン、特にH3を特異的に認識してそれに結合する抗体と接触させる工程を含むことができる。このような抗体の例としては、限定されないが、Upstate Biotechnology, Lake Placid, NYから入手可能な抗アセチル化ヒストンH3が挙げられる。免疫沈降物からのポリヌクレオチドは、その後、免疫沈降物から収集することができる。限定されないが、アダマー、オリゴヌクレオチド、または他の核酸プローブ、および核酸誘導型ヌクレアーゼ（例えば、触媒的に不活性または「死んだ」ヌクレアーゼを含む、Cas9などのCasファミリー酵素）を含む標的特異的な化合物を用いて、同様の標的化濃縮方法を採用する

こともできる。

【0075】

連鎖核酸、例えば、バーコード、パーティション特異的配列、またはパーティション同定配列を有する連鎖核酸は、左ゲノムセグメント、しばしばパーティション特異的配列またはパーティション同定配列（例えば、核酸バーコード）を有する連鎖領域、および右ゲノムセグメントを有する核酸セグメントを生成するように、露出した内部末端に結合させることができ、左ゲノムセグメントおよび右ゲノムセグメントは、供給源細胞において物理的に近接したゲノムセグメントにマッピングされる。

【0076】

露出した核酸末端を結合させる前に、末端を処理することができる。そのような加工は、末端平滑化（*end polishing*）または平滑末端化（*blunt ending*）を含むことができる。平滑末端を露出させた核酸末端は、例えば、他の平滑末端露出核酸末端に直接、またはアダプターもしくはリンカーにライゲーションすることができる。そのような加工は、例えば、テーリング（例えば、Aテーリングまたはアデニル化）によってオーバーハングを生成することを含むことができる。一例では、オーバーハングは1ヌクレオチドのサイズである。一例では、オーバーハングは単一のAヌクレオチドである。尾状の露出した核酸末端は、例えば、他の尾状の露出した核酸末端に直接、またはアダプターもしくはリンカーにライゲーションすることができる。場合によっては、平滑末端化またはテーリングは、ビオチン化核酸などの親和性タグ付け核酸を組み込むことができる。親和性タグは、例えば、下流の捕捉または濃縮工程において使用することができる。他の場合において、平滑末端化またはテーリングは、親和性タグ付け核酸を組み込むことなく（例えば、ビオチン化核酸なしで）実施することができる。その後、親和性タグは、必要に応じて、例えば、アダプターまたはリンカー（例えば、架橋）に付加することができる。一例では、露出した核酸を末端で平滑化し、オーバーハングを生成し、露出した末端を、架橋オリゴを介して結合させる。

【0077】

結合は、ライゲーションなどを介して直接的であり得る。

【0078】

結合は、リンカーまたは架橋を介して、例えば、1つの露出した核酸末端を別の核酸末端に結合する1つ以上のリンカーまたは架橋核酸のライゲーションによるものであり得る。

【0079】

結合は、キャッピング核酸アダプターセグメント、例えば、インテグラーゼまたはトランスポザーゼの取り込みなどのリコンビナーゼの取り込みと一致するものの使用を介するものであり得る。リコンビナーゼ部位を有するアダプターは、露出した核酸末端に付加することができる。次いで、これらの末端は、例えば、組換えによって結合させることができる。

【0080】

一例として *phic31* インテグラーゼバーコード送達を挙げると、細胞同定リンカーまたは細胞特異的リンカー（例えば、核酸バーコード）などのリンカーを、以下のように酵素的に添加することができる。

【0081】

内部核酸末端の露出の後に、インテグラーゼ部位を、露出した核酸末端、例えば、内部末端または露出した線状染色体末端、例えば、テロメアが除去されているものに、ライゲーションすることができる。例示的な組み込み部位は、*attP phic31* インテグラーゼ組み込み部位、または *attP* 組み込み部位を含む核酸であるが、他の組み込み部位は本明細書の開示と一致している。ライゲーションは、核酸断片の集団をもたらす、その少なくともいくつかは個々に、*attP* セグメントを含むセグメントなどの、組み込み部位によって各末端で境界付けられた細胞核酸セグメントを含む。様々な実施形態では、断片化および組み込み部位の結合の一方もしくは両方が、パーティショニングの前に起こ

るか、または断片化および組み込み部位の結合の一方もしくは両方が、パーティショニングの後に起こる。

【0082】

図19A～図19Dは、p h i C 3 1インテグラーゼベースの結合アプローチの例示的な概略図を示す。図19Aでは、インテグラーゼを介したストレプトマイセスへのp h i C 3 1組み込みの概略図が見られる。a t t P（破線によって示される）およびa t t B（実線によって示される）部位を含む核酸が示されているが、様々な実施形態では、a t t Bおよびa t t P以外の部位、ならびにインテグラーゼ以外の酵素活性も企図されており、本明細書の開示と一致する。図19Bでは、インテグラーゼおよび関連タンパク質（丸で示されている）が、細菌ゲノム中のファージa t t B部位およびa t t P配列に結合し、鎖交換を誘発することが分かる。図19Cでは、組み込み事象の結果が見られる。組み込みは、a t t Bおよびa t t Pを欠くが、a t t Bおよびa t t Pの部分のキメラ断片であるa t t Lおよびa t t Rを有する線状核酸に分解される。a t t Lおよびa t t R部位は3 b p短く、a t t Bおよびa t t Pと比較して配列が異なる。図19Dでは、環状の組み込みまたはリンカーゲノムが必要とされないことが分かる。a t t B部位を含有する線状DNAの組み込みにより、a t t P含有DNAの切断が引き起こされる。

10

【0083】

図20は、本明細書で企図される安定化された核酸の露出した内部末端へのインテグラーゼ部位の送達を描いている。例えば、a t t Pは、アダプターライゲーションによって、D N a s e消化クロマチン（円筒形によって示される）の露出した内部末端に送達され得る。核酸は、ヌクレオソームなどの結合部分との接触を保存するように、場合によっては、フェージング情報または三次元の物理的位置を保存するように、安定化させることができる。

20

【0084】

図21は、a t t B部位などのインテグラーゼ部位を使用するリンカー構築物の生成を示す。例えば、最小33ヌクレオチドa t t B部位が組み込みに十分である。フランキング配列は、バーコードまたは特定の供給源（例えば、細胞、液滴、または他のパーティション、生物）由来の核酸を指定する他の配列などの最適な配列を使用して置き換えることができる。図22は、a t t B線状DNAの組み込みによる凝集体内ライゲーションを示す。その結果として、相同（i n p h a s e）するか、または、物理的に近接している核酸セグメントの露出した内部末端を有する線状分子が単一のライブラリー成分に結合している。ライブラリー成分は、無傷の組み込み部位（この場合、a t t P）によって結合されるが、内部組み込み部位は、a t t P関連プライマーがライブラリー断片を増幅し得るように、破壊され、a t t Rおよびa t t L境界に置き換えられている。内部末端隣接配列を得て、それをゲノムまたはコンティグセットにマッピングすることによって、コンティグまたはゲノムセグメントを共通のフェーズまたは細胞内の共通の三次元位置に割り当てることができる。

30

【0085】

図25は、組換えベースの近接ライゲーションプロトコルの別の例を示す。クロスリンクしたクロマチンを含むゲノムDNAは、例えば、D N a s eで消化される。露出した末端を平滑化し、例えば、単一のA塩基オーバーハングでA-ターミングを行う。a t t B部位などのリコンビナーゼ部位を含むAテール適合性アダプターを、露出した末端にライゲーションする。a t t P部位などの対応するリコンビナーゼ部位を有するリンカーを試料に接触させ、リコンビナーゼ酵素（例えば、p h i C 3 1）を用いて組換えを実施することで近接ライゲーションを達成する。リンカーは、下流のプルダウンまたは他の精製もしくはプロセッシングを可能にするために、ビオチン（b）などの親和性剤を任意選択で含む。クロスリンクを脱クロスリンクさせ、例えば、約40 b pのa t t B部位、その後の約150 b pのゲノムDNA領域1、その後a t t R部位および親和性剤を含む約90 b pのリンカー配列、その後約150 b pのゲノムDNA領域2、続いて約40 b pのa t t B部位を含む、近接ライゲーションした核酸を回収する。

40

50

【0086】

図26は例示的なアダプターおよびリンカー配列を用いる、図25に示されるものと同様のプロトコルを示す。最上部には、EPオーバーハングattBアダプター、attB部位GCCCTTGGGCを含む配列GTGCCAGGGCGTGCCCTtGGGGCTCCCCGGGCGCGATC、および逆attB部位GCCCAAGGGCを含む補体配列CGCGCCCCGGGAGCCCCaaGGGCACGCCCTTGGCACを有する、非組換えgDNAが示されている。上から2番目には、attBアダプター、配列GTGCCAGGGCGTGCCCTtGGGGCTCCCCGGGCGCGTGCCCCおよび補体配列GGGGAACGGCGCCCCGGGAGCCCCaaGGGCACGCCCTTGGCACを有する、非組換えgDNAが示されている。上から3番目には、attP部位を含む配列ggagCCCCAAGCTGGGGTAACCTttGAGTTCTCTCAGTTTGGGGaccatggaga/iBiodT/cacCCCCAAGCTGAGAGAACTCaaAGGTTTACCCAGTTTGGGGCACTAC、配列ACCTTTGAGTおよびリンカー配列CATGGAGATCを有する、attP部位およびビオチンを含む非組換えリンカーが示されている。上から4番目には、配列ggagCCCCAAGCTGGGGTAACCTttGAGTTCTCTCAGTTTGGGGaccatggaga/iBiodT/cacCCCCAAGCTGAGAGAACTCaaGGGCACGCCCTTGGCACを有する、attB/gDNAと組換えられたリンカーの一端が示されている。最下部には、配列GCCCTTGAGTを有するattR部位および配列ACTCAAGGGCを有する逆attR部位を含む、配列GTGCCAGGGCGTGCCCCttGAGTTCTCTCAGTTTGGGGaccatggaga/iBiodT/cacCCCCAAGCTGAGAGAACTCaaGGGCACGCCCTTGGCACを有する、attB/gDNAと組換えられたリンカーの両方の末端が示されている。

【0087】

図23Aは、ライブラリー生成を促進するための例示的なリンカー分子およびアダプター分子修飾を示す。リンカー分子には親和性タグ（この場合では、円で示されているビオチン）が付与される一方で、アダプターにはエキソヌクレアーゼ耐性修飾（この場合では、星によって示されているホスホロチオエート化（phosphorothioation）（PS））が与えられる。親和性タグは、リンカー分子が末端隣接分子に組み込まれるかどうかに関係なく、リンカー分子の単離を促進する。リンカー上のエキソヌクレアーゼ耐性修飾は、リンカーが付加されなかった核酸分子の、末端隣接核酸試料分子と統合されなかったリンカー分子への選択的な分解を促進する。図23Bは、タグ付け分子が内部末端に付加されたライゲーション部位に組み込まれるかどうかに関係なく、タグ付け分子の親和性精製（この場合では、半円弧によって示されるストレプトアビジン）を示す。図23Cは、attP部位などのインテグラーゼ部位を保存する親和性単離分子を選択的に増幅するために（この場合、そのような部位をプライマーで標的化することによって）、attP指向性増幅を使用することができることを示している。親和性タグおよびattP部位の存在は、組み込み事象が成功した分子を示す。図23Dは、エキソヌクレアーゼ耐性修飾（この場合、ホスホチオール化（phosphothiolation））を欠く親和性タグ付け分子を除去するためにエキソヌクレアーゼ（扇形または「Pac-Man」によって示される）が使用される代替物を示している。親和性タグおよびエキソヌクレアーゼ耐性部位の存在は、組み込み事象が成功した分子を示す。

【0088】

代替的に、トランスポザーゼ、例えば、Tn3、Tn5、Tn7、またはsleeping beautyトランスポザーゼをバーコード送達に使用することができる。内部核酸末端の露出の後に、モザイク末端を、露出した核酸末端、例えば、内部末端、または露出した線状染色体末端、例えば、テロメアが除去されたものに、ライゲーションすることができる。例示的なモザイク末端は、Tn5モザイク末端、またはTn5モザイク末端を含む核酸であるが、他のモザイク末端は本明細書の開示と一致する。ライゲーションにより核酸断片の集団が生じ、その少なくともいくつかは、Tn5モザイク末端などのモザイ

ク末端によって各末端で境界付けられた細胞核酸セグメントを個々に含んでいる。

【0089】

リコンビナーゼアダプター分子は、配列決定アダプター部位、例えば、P5およびP7部位も含むことができる。図27(上)は、attBアダプター配列GTGCCAGGGCGTGCCCTtGGGCTCCCCGGGCGCG、P7配列GATCGGAAGA GCACACGTCTGAAC TCCAGTCA C、およびP5配列ACACTCTTT C C C T A C A C G A C G C T C T T C C G A T Cを有する、配列決定Yアダプターを有する例示的なattBアダプターを示している。図27(下)は、配列決定アダプターを有するリコンビナーゼアダプターを有する非組換えgDNAの概略図を示す。配列決定アダプターは、組換え後にゲノムDNAと共に残るattB部位の部分に結合され、さらなる増幅またはアダプターライゲーションなしで組換え後の配列決定を可能にする。限定されないが、図27に示されるものを含む配列決定アダプターを含むリコンビナーゼアダプターを使用することで、増幅または別個のアダプター組み込み工程を必要とすることなく、近接ライゲーション産物の直接的な配列決定が可能となる。これにより、得られる配列情報における増幅バイアスなどのバイアスを低減することができる。

10

【0090】

様々な実施形態では、断片化およびモザイク末端の結合の一方もしくは両方が、パーティショニングの前に起こるか、または断片化およびモザイク末端の結合の一方もしくは両方が、パーティショニングの後に起こる。図24は、インテグラーゼ媒介凝集体内ライゲーションを使用する、単細胞HiC(または他の近接ライゲーション技術)のための例示的なシステムを描いている。単細胞核は、インテグラーゼと組み合わせてパーティションの第1のセットに封入される。パーティションは、この場合、エマルジョン中の液滴である。核が鎖切断に供されることで、内部露出末端が生成され、局所的な三次元情報が保存される。アダプターを、露出した内部末端にライゲーションする。アダプターは任意選択でエキソヌクレアーゼ耐性末端を含む。この実施形態では、アダプターは、パーティション識別情報を伝達しない。パーティションの第2のセットでは、固有の分子識別子(UMI)などのパーティション識別配列を有するリンカーが封入され、任意選択で増幅および切断指向性直線化に供される。パーティションの第1および第2のセットは、約1:1の比率で、または2つの細胞からの核酸が単一の結果として生じるパーティションに組み合わせられる可能性が低いような条件下で合体する。

20

30

【0091】

インテグラーゼ部位またはモザイク末端などのリコンビナーゼ部位は、場合によっては、内部核酸末端上にライゲーションされる未修飾の一本鎖断片または二本鎖断片上で担持され得る。代替的に、その後の配列決定ライブラリーのクリーンアップを容易にするために、attP配列またはモザイク末端、例えば、Tn3、Tn5、Tn7、またはsleeping beautyトランスポザーゼモザイク末端などの組み込み部位を有するいくつかの一本鎖または二本鎖断片は、エキソヌクレアーゼまたは他の核酸分解活性を妨害する修飾などの少なくとも1つの修飾を含むことができる。例としては、組み込み部位を有する二本鎖断片が各末端に付加された断片のエキソヌクレアーゼ分解を排除するためのチオ硫酸修飾が挙げられる。

40

【0092】

しばしば、組み込み部位またはモザイク末端などのリコンビナーゼ部位は、そのような組み込み部位またはモザイク末端の配列、例えば、attP配列もしくはTn3、Tn5、Tn7、またはsleeping beautyトランスポザーゼモザイク末端が、隣接する核酸の細胞源を指定するためには使用されないという点で、非特異的である。代替的に、しばしば核酸パーティショニングの後に、パーティションに、組み込み部位またはモザイク末端に隣接する別個の、特異的な配列または細胞識別配列(例えば、核酸バーコード)を有するアダプターを提供するか、または別個の組み込み部位またはモザイク末端を提供することができ、したがって、第1のパーティションの核酸は、第1の同定セグメントを有する組み込みセグメントまたはモザイク末端を受け取り、第2のパーティション

50

の核酸セグメントは、第2の同定セグメントを有する組み込みセグメントを受け取るようになる。

【0093】

次いで、インテグラーゼ *attP* セグメントを含む境界などのリコンビナーゼ境界を有する断片を、共通の溶液中で、*attB* *phiC31* 組み込み部位などの組み込み部位に接触させることができる。例えば、組み込み酵素は *phi31* インテグラーゼを含むことができ、組み込み境界は *attP* セグメントを含むことができ、組み込み部位は *attB* 組み込み部位を含むことができる。代替的に、断片は、*Tn3*、*Tn5*、*Tn7* などのモザイク末端境界、または *sleeping beauty* トランスポザーゼモザイク末端境界を有する。

10

【0094】

attB 組み込み部位、または *Tn3*、*Tn5*、*Tn7*、または *sleeping beauty* トランスポザーゼモザイク末端などのリコンビナーゼ部位が、セグメントまたは細胞源（例えば、核酸バーコード）に特異的なものなどの、パーティションまたは細胞を同定する配列を有する連鎖セグメントに隣接する場合、その配列は、特定のまたは共通の細胞源またはパーティションから生じるものとして隣接細胞核酸を同定し、したがって、共通の細胞識別セグメントまたはパーティション識別セグメントによって結合された共通の細胞からの複数の露出した末端は、配列決定の前にまたは配列決定と同時に、第2のパーティションの断片とともにバルク化される場合であっても、共通の細胞から生じるものとして容易に同定することができるようになる。

20

【0095】

細胞識別配列がリコンビナーゼ部位境界断片を介して送達される場合、組み込みまたは転位は好ましくは、パーティショニングの後に行われる。少なくともいくつかのパーティションの核酸内容物は、それによって、そのリンカーの細胞識別配列によって識別することができ、したがって、複数の細胞源に由来する核酸が配列決定のためにバルク化された後であっても、内部末端対と、大部分または完全に配列決定されたゲノムを最大限含むコンティグセットにおいてそれらがマッピングされる近傍に割り当てられる近接情報とを、試料の少なくとも1つの他の細胞と区別される共通の細胞に対して割り当てることができる。そうすることで、予測される核酸三次元立体構造の差異を確立することができるようになる。

30

【0096】

組換え部位境界断片は、任意選択で細胞またはパーティション指定配列（例えば、核酸バーコード）を含むリンカー領域によって結合された左境界断片および右境界断片（例えば、*attB* 部位または *Tn3*、*Tn5*、*Tn7*、または *sleeping beauty* トランスポザーゼモザイク末端）を様々に含む。リンカー領域は任意選択で、その後の単離を促進するための部分をさらに含む。多くの親和性タグまたは修飾塩基は、本明細書の開示に一致している。例示的な部分は、インテグラーゼまたはトランスポザーゼ処理後のリンカーの物理的または化学的な単離を促進する。アビジンまたはストレプトアビジンベースの単離を促進し得る1つまたは複数のビオチンタグなどの、任意の数の親和性タグが本明細書の開示と一致する。代替的に、インテグラーゼまたはトランスポザーゼ活性を妨害することなく単離を促進する任意の抗原、受容体、またはリガンドが、本明細書のいくつかの実施形態に適している。

40

【0097】

上述のように、いくつかのライブラリー生成アプローチは、取り込まれていない試薬を選択的に除去する工程などのクリーンアップ工程を含む。例えば、エキソヌクレアーゼ処理は、結合していないリンカー分子、いかなる組み込み部位も結合していないゲノム断片、または結合していないリンカー分子およびいかなる組み込み部位も結合していないゲノム断片の両方を選択的に除去するためにしばしば使用される。チオ硫酸骨格などのエキソヌクレアーゼ耐性修飾を有する組み込み部位断片にライゲーションされたゲノム断片は、その末端からエキソヌクレアーゼ分解に耐性があり、チオ硫酸骨格などのエキソヌクレア

50

ーゼ耐性修飾を有する組み込み部位断片によって両方の末端で結合された核酸分子は、両方の末端で分解に耐性があり、エキソヌクレアーゼ処理を生き残ることができる。

【0098】

代替的にまたは組み合わせて、いくつかのリンカー分子は、組換え部位、例えば、attP組み込み部位、またはTn3、Tn5、Tn7、またはsleeping beautyトランスポザーゼモザイク末端の反対側に逆親和性タグを含み、逆親和性タグは、組換え反応の成功に従って除去される。そのような場合、望ましくない試薬は、逆親和性タグの結合パートナーと接触させることによって除去することができる。

【0099】

インテグラーゼ活性は、組み込み事象の一部として、attB部位およびattP部位などの両方の組み込み部位を部分的に破壊する。したがって、attP組み込み部位などのライゲーションされたアダプター部位にアニーリングするプライマーを、単独で、またはリンカーベースの単離と組み合わせて、設計することによって、細胞またはアリコート識別情報および内部末端隣接情報が増幅され、場合によっては、配列決定または他の下流分析を容易にするように、少なくとも1つのリンカーにまたがるクローンアンプリコンを生成することができる。

【0100】

ライブラリーの生成および任意選択でライブラリーのクリーンアップの後に、核酸を完全にまたは部分的に配列決定して、細胞識別または細胞特異的三次元核酸位置評価に十分な情報を得ることができる。上述のように、配列決定は好ましくは、ライブラリー成分の各ゲノム末端をそのゲノム遺伝子座にマッピングするのに十分な少なくともいくつかのゲノム配列を取得し、さらに、正確なまたは可能性のある起始細胞を同定するのに十分な連鎖配列を取得するように行われる。したがって、互いに物理的に近接しているゲノムの2つの領域の有用な配列情報を得ると同時に、この物理的な立体構造が生じる細胞の有用な情報も得て、それを、その細胞において同時に生じる他の物理的な立体構造情報の文脈で評価することができるようになる。多くの場合、この情報は完全長配列決定ではなくペアエンド配列決定によって得られるが、両方のアプローチおよび他のアプローチは本明細書の開示と一致している。

【0101】

第2の細胞上の物理的な立体構造と識別可能な単細胞などの細胞における核酸の物理的な立体構造の決定に関する組成物および方法は、本明細書の開示と一致する多くのシステム上で実施することができる。いくつかのシステムは、エマルジョンの第1の液滴中、またはウェル中、例えば、ウェルプレート上での固定細胞核酸材料の分配を含む。これらの液滴は、本明細書に記載されるように任意選択でエキソヌクレアーゼ耐性となるように任意に修飾されたインテグラーゼ部位またはモザイク末端などのリコンビナーゼ部位、ならびにインテグラーゼまたはトランスポザーゼ酵素およびリガーゼ酵素を含む。別個に、リンカー核酸分子は、エマルジョンの第1の液滴への送達のために構成され得る。リンカー核酸は任意選択で、第2のエマルジョンまたは第2のウェルの液滴に分配され、任意選択で、例えば、ローリングサークル増幅を使用して増幅され、所定のリンカー分子／エマルジョン液滴の複数のコピーを生成するように処理されることができる。

【0102】

次いで、第2のエマルジョン液滴および第1のエマルジョン液滴は、インテグラーゼまたはトランスポザーゼ適合性リンカーを用いてインテグラーゼまたはトランスポザーゼライゲーション核酸断片を組み立てるために対で合体され得、しばしば、液滴あたり均一な標識を示す。しかしながら、核酸試料あたり2つ以上の識別子を有する液滴は、とりわけ、データ分析が液滴中の複数の種類のタグの存在を示す場合、依然として有意義なデータをもたらすことが可能であり得る。

【0103】

対合体の代替物として、場合によっては、インテグラーゼまたはトランスポザーゼ適合性リンカーは、全体として参照により本明細書に組み込まれる、2017年11月23日

10

20

30

40

50

に公開されたUS20170335369A1に記載されるものなど、液滴からストリームへの合体を介して第1のエマルジョン液滴と接触する試薬ストリーム中の固体粒子のコロニーとして送達することができる。リンカー核酸は、固体粒子上またはゲル中で、任意選択で増幅することができる。第1のエマルジョン液滴は、ストリームに合体させることができ、第2のエマルジョン液滴は、ストリームをセグメント化または分割することによって回収することができ、1:1、1:1超、または1:1未満などの所望の割合の核酸クラスター対リンカー粒子が得られるようになる。

【0104】

代替的に、いくつかのシステムおよび方法は、固定された細胞核酸材料をチップまたはプレートのウェルに分配すること、その後、未増幅のまたは上記のように増幅されたリンカー核酸をパーティションに送達することを含む。

10

【0105】

代替的に、場合によっては、リンカー核酸の送達は、分割から時間的に分離されない。むしろ、リンカー核酸または酵素活性または酵素活性に必要な因子は、リンカーを介するなどしてリンカーの核酸試料露出末端への共有結合をもたらす酵素活性を一時的に活性化するように、熱、電磁活性化、または他の投与などの特定の処置まで隔離される。

【0106】

多くの組み込み酵素は、本明細書の開示と一致する。ThermoFisherによって市販されているものなどのPhic31インテグラーゼは、方法の実施、システムの操作、および本明細書の組成物における使用について多くの利点を示す。このインテグラーゼのいくつかの利点は以下の通りである。これは、小さな組み込み部位(attB/attP)を使用する。酵素自体は小さな単一ポリペプチドである。組み込みは、組み込み事象を切除するための別個の酵素を使用することなく不可逆的である。活性は高く、酵素は容易に操作されて活性を変化させる。それにもかかわらず、多くの組み込みシステムが本明細書の開示と一致しているように、その使用は他の酵素の除外に必要ではない。本開示の態様はPhic31インテグラーゼに関して記載され得るが、任意の適合性のある酵素の使用が企図される。

20

【0107】

多くのトランスポザーゼ酵素は、本明細書の開示と一致する。Lucigenによって市販されているものなどのTn5トランスポザーゼは、本方法の実施、システムの操作、および本明細書の組成物における使用について多くの利点を示す。このトランスポザーゼのいくつかの利点は以下の通りである。Tn5は19bpのモザイク末端認識配列を使用し、挿入はほとんど偏りがなく安定しており、Tn5はインビボ転位のために細胞に送達され得、またはインビトロ反応のために単離核酸に送達され得る。それにもかかわらず、Tn3、Tn7、またはsleeping beautyトランスポザーゼなどの多くのトランスポザーゼ系は本明細書の開示と一致するため、その使用は他の酵素の除外に必要ではない。本開示の態様は、Tn3、Tn5、Tn7、またはsleeping beautyトランスポザーゼに関して記載され得るが、任意の適合性のある酵素の使用が企図される。

30

【0108】

ライブラリー成分から得られる配列情報は、Hi-C、Chicago（登録商標）インビトロ近接ライゲーション、または他の三次元立体構造解析の文脈におけるものなどの多くのアプローチによって評価される。重要なことに、ゲノムまたは特定のコンティグの特定の領域への末端隣接配列マッピングの頻度を細胞特異的に評価することができるように、細胞特異的リード対の頻度を得ることができる。すなわち、可能性のある三次元立体構造の細胞特異的な発生を評価することができる。場合によっては、三次元立体構造における細胞特異的な距離と相関するシグナルの細胞特異的な強度を評価することもでき、その結果、核酸の特定の領域が、1つの細胞において、それらが同等であるが「弱い」またはより遠くに近接している第2の細胞に対して比較的近接しており、第3の細胞にある間、近接を示すシグナルがないと結論付けることができる。すなわち、三次元立体構造の定

40

50

性的評価および定量的評価の両方は、本明細書の開示と一致する。場合によっては、第2の領域に対する1つの領域の近接は、第2のクラスターのクラスター成分との対をなすペアエンドリードにおいて同時に生じる第1のクラスターのクラスター成分の数を数えることによって、とりわけ、固有のパーティションタグなどの共通のパーティション識別配列を共有するライブラリー成分において、少なくとも部分的に評価される。

【0109】

構成情報は、複数のライブラリー成分における同一の末端隣接配列の複数回の発生を通して作製される必要はない。むしろ、場合によっては、第2の末端隣接配列マッピング部位付近に（共通の「クラスター」に）マッピングされる末端隣接配列は、クラスターの両方のメンバーが、ゲノムなどの核酸参照の第2の領域上の第2のクラスターの非同一領域にマッピングされる場合に、三次元立体構造評価を強化することができる。

10

【0110】

場合によっては、本明細書に開示される方法は、ポリヌクレオチドまたはその配列セグメントを標識および／または会合させ、そのデータを様々な用途に利用するために使用される。場合によっては、本開示は、約10,000未満、約20,000未満、約50,000未満、約100,000未満、約200,000未満、約500,000未満、約100万未満、約200万未満、約500万未満、約1000万未満、約2000万未満、約3000万未満、約4000万未満、約5000万未満、約6000万未満、約7000万未満、約8000万未満、約9000万未満、約1億未満、約2億未満、約3億未満、約4億未満、約5億未満、約6億未満、約7億未満、約8億未満、約9億未満、または約10億未満のリード対を有する、高度に連続したかつ正確なヒトゲノムアセンブリを生成する方法を提供する。場合によっては、本開示は、約50%、約60%、約70%、約75%、約80%、約85%、約90%、約91%、約92%、約93%、約94%、約95%、約96%、約97%、約98%、約99%、またはそれ以上の精度で、ヒトゲノム中のヘテロ接合性バリエーションの約50%、約60%、約70%、約75%、約80%、約85%、約90%、約91%、約92%、約93%、約94%、約95%、約96%、約97%、約98%、約99%、またはそれ以上をフェージングするか、それらに物理的連鎖情報を割り当てる方法を提供する。

20

【0111】

いくつかの実施形態では、本明細書に記載される組成物および方法は、メタゲノム、例えば、ヒトの腸で見られるものの調査を可能にする。したがって、所定の生態学的環境に生息するいくつかまたはすべての生物の部分的または完全なゲノム配列を調査することができる。例としては、すべての腸内微生物、皮膚の特定の領域で見られる微生物、および毒性廃棄物部位に生存する微生物のランダムな配列決定が挙げられる。これらの環境における微生物集団の組成は、本明細書に記載される組成物および方法、ならびにそれらのそれぞれのゲノムによってコードされる相互に関連する生化学の態様を使用して、決定することができる。本明細書に記載の方法は、複雑な生物学的環境、例えば、2、3、4、5、6、7、8、9、10、12、15、20、25、30、40、50、60、70、80、90、100、125、150、175、200、250、300、400、500、600、700、800、900、1000、5000、10000またはそれ以上を超える生物および／または生物のバリエーションを含む生物学的環境からのメタゲノム研究を可能にすることができる。

30

40

【0112】

したがって、本明細書に開示される方法は、無傷のヒトゲノムDNA試料に適用され得るが、逆転写RNA試料、循環遊離DNA試料、癌組織試料、犯罪現場試料、古細菌試料、非ヒトゲノム試料などの広範で多様な核酸試料、または1を超える生物、例えば、実験室条件下で容易に培養されない生物からの遺伝情報を含む環境試料などの環境試料にも適用され得る。

【0113】

癌ゲノム配列決定に必要とされる高い精度は、本明細書に記載される方法およびシステ

50

ムを使用して達成することができる。不正確な参照ゲノムは、癌ゲノムを配列決定する際に塩基呼び出しを困難なものにしかねない。不均一な試料および小さな出発物質、例えば、生検により得られた試料がさらなる困難をもたらす。さらに、大規模な構造的なバリエーションおよび／またはヘテロ接合性の損失の検出はしばしば、癌ゲノム配列決定の他、体細胞バリエーションと塩基呼び出しのエラーとを区別する能力にも、必要不可欠である。

【0114】

本明細書に記載されるシステムおよび方法は、2、3、4、5、6、7、8、9、10、12、15、20、またはそれ以上の様々なゲノムを含む複合試料から正確な長い配列を生成し得る。通常の、良性の、および／または腫瘍由来の混合試料は任意選択で、通常の対照を必要とすることなく、解析されてもよい。いくつかの実施形態において、100 ng または数百ほどのゲノム当量の小さな出発試料は、正確な長い配列を生成するために利用される。本明細書に記載されるシステムおよび方法は、大規模な構造バリエーションおよび再編成の検出が可能となる場合があり、フェージングされたバリエーションコールは、約1 kbp、約2 kbp、約5 kbp、約10 kbp、約20 kbp、約50 kbp、約100 kbp、約200 kbp、約500 kbp、約1 Mbp、約2 Mbp、約5 Mbp、約10 Mbp、約20 Mbp、約50 Mbp、または約100 Mbp、またはそれ以上のヌクレオチドにまたがる長い配列にわたって得られ得る。例えば、フェージングされたバリエーションコールは、約1 Mbp または約2 Mbp にまたがる長い配列にわたって得られる得る。

【0115】

特定の態様において、本明細書に開示される方法は、単一のDNA分子に由来する複数のコンティグを組み立てるために使用される。場合によっては、方法は、複数のナノ粒子にクロスリンクされる単一のDNA分子から複数のリード対を生成する工程と、リード対を使用してコンティグを組み立てる工程とを含む。特定の場において、単一のDNA分子は、細胞の外側でクロスリンクされる。場合によっては、リード対の少なくとも0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、35%、40%、45%、または50%は、単一のDNA分子上の1 kB、2 kB、3 kB、4 kB、5 kB、6 kB、7 kB、8 kB、9 kB、10 kB、15 kB、20 kB、30 kB、40 kB、50 kB、60 kB、70 kB、80 kB、90 kB、100 kB、150 kB、200 kB、250 kB、300 kB、400 kB、500 kB、600 kB、700 kB、800 kB、900 kB、または1 MB よりも大きい距離に及ぶ。特定の場において、リード対の少なくとも0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、または20%は、単一のDNA分子上の5 kB、6 kB、7 kB、8 kB、9 kB、10 kB、15 kB、20 kB、30 kB、40 kB、50 kB、60 kB、70 kB、80 kB、90 kB、100 kB、150 kB、または200 kB より大きい距離に及ぶ。さらなる場において、リード対の少なくとも0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、または5%は、単一のDNA分子上で20 kB、30 kB、40 kB、50 kB、60 kB、70 kB、80 kB、90 kB、または100 kB を超える距離に及ぶ。特定の場において、リード対の少なくとも1%または5%は、単一のDNA分子上で50 kB または100 kB よりも大きい距離に及ぶ。場合によっては、リード対は、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、または60 日以内に生成される。特定の場において、リード対は、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、または18 日以内に生成される。さらなる場において、リード対は、7、8、9、10、11、12、13、または14 日以内に生成される。特定の場において、リード対は、7 または14 日以内に生成される。

【0116】

本明細書に記載される方法およびシステムを使用して決定されるハプロタイプは、コンピュータリソース、例えば、クラウドシステムなどのネットワーク上のコンピュータリソースに割り当てられてもよい。短いバリエーションコールは、必要であれば、コンピュータリソースに保管される関連情報を使用して修正することができる。構造バリエーションは、短いバリエーションコールからの組み合わせられた情報、およびコンピュータリソースに保管された情報に基づいて検出することができる。ゲノムの問題となる部分、例えば、セグメント重複、構造多型 (structural variation) の傾向がある領域、高度に可変的でありかつ医学的に関連するMHC領域、セントロメアおよびテロメアの領域、ならびに、反復領域、低い配列精度、高いバリエーション率、ALU反復、セグメント重複を持つものを含む他の異質染色質領域、または当該技術分野での他の関連する問題となる部分を、精度の向上ために再度組み立てることができる。

10

【0117】

試料の種類は、局所的に、またはクラウドなどのネットワーク接続されたコンピュータリソースにおいて、配列情報に割り当てることができる。情報のソースが分かっている場合、例えば、情報のソースが癌または正常な組織由来である場合、このソースは試料の種類の一部として試料に割り当てることができる。他の試料の種類は通常、限定されないが、組織の種類、試料収集方法、感染の存在、感染の種類、処理方法、試料のサイズなどを含む。癌ゲノムとの比較における通常のゲノムのなどの、完全または部分的な比較ゲノム配列が利用可能である場合、試料データと比較ゲノム配列との差を判定することができる、任意選択で出力することができる。

20

【0118】

本開示の方法は、選択的なゲノムの関心領域の他に、選択的な関心領域と相互作用し得るゲノム領域の遺伝子情報の分析に使用され得る。本明細書に開示されるような増幅方法は、限定されないが、米国特許第6,449,562号、第6,287,766号、第7,361,468号、第7,414,117号、第6,225,109号、および第6,110,709号でみられるものなど、遺伝子分析のためのデバイス、キット、および方法に使用することができる。場合によっては、本開示の増幅方法は、多型の存在の有無を判定するためのDNAハイブリダイゼーション研究のために標的核酸を増幅するために使用することができる。多型または対立遺伝子は、遺伝病などの疾患または疾病に関連付けることができる。いくつかの他の場合において、多型は、疾患または疾病に対する感受性に関連付けることができ、例えば、多型は、中毒、退行性および加齢性の疾病、癌などに関連付けられる。他の場合において、多型は、冠動脈の健康の増大といった有用な特色、HIVもしくはマラリアなどの疾患に対する抵抗性、または骨粗鬆症、アルツハイマー病、もしくは痴呆などの成人病に対する抵抗性に関連付けることができる。

30

【0119】

本開示の組成物および方法は、診断、予後、治療、患者の層別化、薬物の開発、処置の選択、およびスクリーニングの目的のために使用することができる。本開示は、本開示の方法を使用して多くの様々な標的分子が単一の生体分子試料から一度に分析することができるという利点を提供する。これにより、例えば、様々な診断試験を1つの試料上で行うことが可能となる。

40

【0120】

本明細書において提供される方法は、これらの反復領域によってもたらされる実質的な障壁を克服することによってゲノムの分野を大幅に進歩させることができ、それによってゲノム分析の多くの領域における重要な進歩を可能にすることができる。以前の技術を用いてde novoアセンブリを実施するためには、多くの小さな足場に断片化されたアセンブリを準備するか、または、大きな挿入ライブラリーを生成するための、もしくはより連続したアセンブリを生成するための他のアプローチを使用するために、かなりの時間および資源を必要とするかのいずれかでなければならない。そのようなアプローチは、非常に深いシーケンシングカバレッジの獲得、BACまたはホスミドライブラリーの構築、

50

光学マッピング、または、おそらく、これらおよび他の技術のいくつかの組み合わせを含み得る。資源および時間の厳しい要件のせいで、そのようなアプローチが大部分の小規模研究室に浸透せず、非モデル生物の研究が妨げられている。本明細書に記載される方法は、非常にロングレンジのリードセットを生成することができるため、デノボアセンブリは、単一の配列決定の実行で達成され得る。これにより、組立コストが桁違いに削減され、必要とされる時間が数ヶ月または数年から数週間にまで短縮される。場合によっては、本明細書に開示される方法は、14日未満、13日未満、12日未満、11日未満、10日未満、9日未満、8日未満、7日未満、6日未満、5日未満、4日未満、3日未満、2日未満、1日未満、または前述の指定された期間のいずれか2つの間の範囲内で、複数のリードセットを生成することを可能にする。場合によっては、方法は、約10日~14日で複数のリードセットを生成することを可能にする。生物の大部分のニッチについてさえゲノムを構築することは日常的となり、系統発生分析は比較の欠如に悩まされることなく、Genome 10kなどのプロジェクトを実現することができる。

10

【0121】

本明細書に記載される方法は、以前に提供されたコンティグ情報、以前に生成されたコンティグ情報、またはデノボ合成されたコンティグ情報を、染色体またはより短い連続する核酸分子などの物理的な連鎖群に割り当てることを可能にする。同様に、本明細書に開示される方法は、前記コンティグが物理的な核酸分子に沿って直線的な順序で互いに対して配置されることを可能にする。同様に、本明細書に開示される方法は、前記コンティグが物理的な核酸分子に沿って直線的な順序で互いに対して配向されることを可能にする。

20

【0122】

同様に、本明細書に開示される方法は、医療目的のための構造解析およびフェージング解析における進歩を提供することができる。癌、同じ種類の癌を持つ個体、または同じ腫瘍の中でさえ、驚くべき異種性が存在する。結果として生じる効果から原因となるものを引き出すことは、試料あたりの低いコストで非常に高い精度およびスループットを必要とする。オーダーメイド医療の領域では、ゲノムケアの絶対的基準のうち1つは、大小の構造再編成および新規な突然変異を含む、徹底的に特徴付けられかつフェージングされたすべてのバリエーションを有する、配列決定されたゲノムである。以前の技術でこれを達成するためには、現在ではあまりに高価でありかつ慣例的な医療処置を要求する、デノボアセンブリに必要なものと同様の労力を必要とする。場合によっては、本明細書に開示される方法は、低コストで完全に正確なゲノムを迅速に産生し、それによって、ヒト疾患の研究および処置において多くの高度に求められる能力をもたらす。

30

【0123】

さらに、本明細書に開示される方法をフェージングに適用することは、統計的アプローチの利便性を、家族分析の精度と組み合わせることができ、いずれかの方法を単独で使用するものよりも大きな節約—費用、労力、および試料—をもたらす。以前の技術では禁止的である非常に望ましいフェージング分析である新規バリエーションフェージング解析は、本明細書に開示される方法を使用して容易に実施することができる。このことは、ヒトの変種の大部分が稀であるため（5%未満のマイナー対立遺伝子頻度）、特に重要である。フェージング情報は、非連鎖遺伝子型と比較して、高度に結合したハプロタイプのネットワーク（単一染色体に割り当てられたバリエーションの集合）から有意な利点を得る集団遺伝子研究にとって貴重である。ハプロタイプ情報は、集団のサイズ、移動、および亜集団間の交換の歴史的变化のより高解像度の研究を可能にし得、特定のバリエーションを特定の親および祖父母に遡って追跡することを可能にする。これは、ひいては、疾患に関連するバリエーションの遺伝的伝達、および単一の個体においてまとめられた場合のバリエーション間の相互作用を明らかにする。さらなる場合において、本開示の方法は、極端にロングレンジのリードセット（XLR S）または極端にロングレンジのリード対（XLR P）ライブラリーの調製、配列決定、および分析を可能にする。

40

【0124】

本開示のいくつかの実施形態では、対象由来の組織またはDNA試料が提供され、本方

50

法は、組み立てられたゲノム、コールされたバリエーション（大きな構造バリエーションを含む）とのアラインメント、フェージングされたバリエーションコール、または任意の追加の解析に戻る。他の実施形態において、本明細書に開示される方法は、個体のためのXLRPライブラリーを直接提供する。

【0125】

様々な実施形態では、本明細書に開示される方法は、大きい距離によって分離される極端にロングレンジのリード対を生成する。この距離の上限は、大きなサイズのDNA試料を集める能力により改善され得る。場合によっては、リード対は、最大50、60、70、80、90、100、125、150、175、200、225、250、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000 kbp、またはそれ以上のゲノム距離におよぶ。場合によっては、リード対は、最大500 kbpのゲノム距離におよぶ。他の場合には、リード対は、最大2000 kbpのゲノム距離におよぶ。本明細書に開示される方法は、分子生物学の標準技術に基づいて組み込みかつ構築することができ、さらに効率、特異性、およびゲノムカバレッジの増大に十分に適している。場合によっては、リード対は、約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、60、または90日未満で生成される。場合によっては、リード対は約14日未満で生成される。さらなる場合には、リード対は約10日未満で生成される。場合によっては、本開示の方法は、複数のコンティグの正確な順序付けおよび／または配向における少なくとも約50%、約60%、約70%、約80%、約90%、約95%、約99%、または約100%の精度で、リード対の約5%、約10%、約15%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%、約95%、約99%、または約100%超を提供する。場合によっては、方法は、複数のコンティグの正確な順序付けおよび／または配向において約90~100%の精度を提供する。

10

20

【0126】

他の実施形態において、本明細書に開示される方法は、現在利用される配列決定技術と共に使用される。場合によっては、この方法は、十分に試験されおよび／または広く配備された配列決定機器と組み合わせて使用される。さらなる実施形態において、本明細書に開示される方法は、現在利用される配列決定技術に由来する技術と手法と共に使用される。

30

【0127】

本明細書に開示される方法は、多種多様な生物のためのデノボゲノムアセンブリを劇的に単純化することができる。以前の技術を使用して、そのようなアセンブリは、現在、経済的なメイトペアライブラリーの短い挿入物によって制限されている。フォスミドでアクセス可能な最大40~50 kbpのゲノム距離でリード対を生成することが可能であり得るが、これらは高価であり、扱いにくく、ヒトでは300 kbp~5 Mbpのサイズの範囲であるセントロメア内のものを含む最も長い反復ストレッチに及ぶには短すぎる。場合によっては、本明細書に開示される方法は、長い距離（例えば、メガベース以上）に及ぶことが可能なリード対を提供し、それによって、これらの足場完全性の課題を克服する。したがって、染色体レベルのアセンブリの産生は、本明細書に開示される方法を利用することによって日常的であり得る。同様に、ロングレンジフェージング情報の獲得は、集団のゲノム研究、系統発生研究、および疾患研究にさらなる膨大な力を提供することができる。特定の場合において、本明細書に開示される方法は、多数の個体についての正確なフェージングを可能にし、したがって、集団および深い時間レベルでゲノムをプローブする我々の能力の幅および深さを拡張する。

40

【0128】

個別化医療の分野では、本明細書に開示される方法から生成されるXLR Sリードセットは、正確で、低コストで、フェージングされた、迅速に産生される個人ゲノムへの有意義な進歩を表している。以前の方法は、長距離でバリエーションをフェージングする能力が不

50

十分であり、それによって、化合物ヘテロ接合遺伝子型の表現型の影響の特徴付けを妨げる。さらに、ゲノム疾患に実質的な関心対象の構造バリエーションは、それらを研究するために使用されるリードおよびリード挿入物と比較してサイズが大きいいため、以前の技術を用いて正確に同定および特徴付けすることが困難である。数十キロベース～メガベース以上にわたるリードセットは、この困難さを軽減するのに役立つことができ、それによって、構造多型の高度に並列かつ個別化された分析を可能にする。

【0129】

基本的な進化研究および生物医学研究は、ハイスループット配列決定における技術的進歩によって推進され得る。現在、大量のDNA配列データを生成することは比較的安価である。しかしながら、以前の技術を用いて高品質で高度に連続したゲノム配列を作製することは、理論上および実際には困難である。さらに、ヒトを含む多くの生物は二倍体であり、各個体はゲノムの2つの半数体コピーを有する。ヘテロ接合性の部位（例えば、母親によって与えられる対立遺伝子が父親によって与えられる対立遺伝子と異なる場合）では、どの対立遺伝子のセットがどの親から来たかを知ることは困難である（ハプロタイプフェージングとして知られる）。この情報は、疾患および形質の関連研究などのいくつかの進化的研究および生物医学的研究を実施するために極めて重要であり得る。

【0130】

本開示は、所定のゲノム内のクロマチン複合体などの複合体に結合した単一の物理的核酸分子からの配列リードに対応する、短い、中程度の、および長い結合のハイスループット発見のために、タグ付き配列リードとDNA調製のための技術を組み合わせるゲノムアセンブリのための方法を提供する。本開示はさらに、ゲノムアセンブリを補助するために、ハプロタイプフェージングのために、および／またはメタゲノム研究のために、これらの結合を使用する方法を提供する。本明細書に提示される方法は、対象のゲノムのアセンブリを決定するために使用することができるが、特定の場合において、本明細書に提示される方法は、染色体などの対象のゲノムの部分のアセンブリ、または様々な長さの対象のクロマチンのアセンブリを決定するために使用されることも理解されたい。特定の場合において、本明細書に提示される方法は、非染色体核酸分子のアセンブリを決定または指示するために使用されることも理解されたい。実際に、非反復コンティグを分離する反復領域の存在によって配列決定が複雑である任意の核酸は、本明細書に開示される方法を使用して促進され得る。

【0131】

さらなる場合において、本明細書に開示される方法は、少量の材料を用いた遺伝子型アセンブリ、ハプロタイプフェージング、およびメタゲノミクスについての正確かつ予測的な結果を可能にする。場合によっては、約100ピコグラム（pg）未満、約200pg未満、約300pg未満、約400pg未満、約500pg未満、約600pg未満、約700pg未満、約800pg未満、約900pg未満、約1.0ナノグラム（ng）未満、約2.0ng未満、約3.0ng未満、約4.0ng未満、約5.0ng未満、約6.0ng未満、約7.0ng未満、約8.0ng未満、約9.0ng未満、約10ng未満、約15ng未満、約20ng未満、約30ng未満、約40ng未満、約50ng未満、約60ng未満、約70ng未満、約80ng未満、約90ng未満、約100ng未満、約200ng未満、約300ng未満、約400ng未満、約500ng未満、約600ng未満、約700ng未満、約800ng未満、約900ng未満、約1.0マイクログラム（μg）未満、約1.2μg未満、約1.4μg未満、約1.6μg未満、約1.8μg未満、約2.0μg未満、約2.5μg未満、約3.0μg未満、約3.5μg未満、約4.0μg未満、約4.5μg未満、約5.0μg未満、約6.0μg未満、約7.0μg未満、約8.0μg未満、約9.0μg未満、約10μg未満、約15μg未満、約20μg未満、約30μg未満、約40μg未満、約50μg未満、約60μg未満、約70μg未満、約80μg未満、約90μg未満、約100μg未満、約150μg未満、約200μg未満、約300μg未満、約400μg未満、約500μg未満、約600μg未満、約700μg未満、約800μg未満、約900μg未満、また

は約1000 μ g未満のDNAが、本明細書に開示される方法とともに使用される。場合によっては、本明細書に開示される方法で使用されるDNAは、約10,000,000未満、約5,000,000未満、約4,000,000未満、約3,000,000未満、約2,000,000未満、約1,000,000未満、約500,000未満、約200,000未満、約100,000未満、約50,000未満、約20,000未満、約10,000未満、約5,000未満、約2,000未満、約1,000未満、約500未満、約200未満、約100未満、約50未満、約20未満、または約10未満の細胞から抽出される。

【0132】

二倍体ゲノムでは、染色体対上の相同位置にマッピングするのではなく、どの対立遺伝子バリエーションが同じ染色体上で物理的に連合されているかを知ることがしばしば重要である。対立遺伝子または他の配列を二倍体染色体対の特定の物理的染色体にマッピングすることは、ハプロタイプフェージングとして公知である。ハイスループット配列データからのショートリードは、とりわけ、最も頻繁にあることだが、対立遺伝子バリエーションが最長の単一リードよりも長い距離だけ離れている場合に、どの対立遺伝子バリエーションが結合されているかを直接観察することを可能にすることはほとんどない。ハプロタイプフェージングのコンピュータによる推論は、長距離では信頼できない可能性がある。本明細書に開示される方法は、リード対上の対立遺伝子バリエーションを使用して、どの対立遺伝子バリエーションが物理的に結合されるかを決定することを可能にする。

【0133】

様々な場合において、本開示の方法および組成物は、複数の対立遺伝子バリエーションに関して、二倍体または多倍体のゲノムのハプロタイプフェージングを可能にする。したがって、本明細書に記載される方法は、標識された配列セグメントからのバリエーション情報および／またはそれを使用するアセンブルされたコンティグに基づいて、結合された対立遺伝子バリエーションの決定を提供する。対立遺伝子バリエーションの例としては、限定されないが、1000ゲノム、UK10K、HapMap、およびヒト間の遺伝的変異を発見するための他のプロジェクトから知られているものが挙げられる。場合によっては、例えば、シャルコー・マリー・トゥース神経障害 (Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. N. Engl. J. Med. 362:1181-91, 2010) を引き起こすSH3TC2の両方のコピーにおける非連鎖の不活性化突然変異、および高コレステロール血症9 (Rios J, Stein E, Shendure J, et al. Hum. Mol. Genet. 19:4313-18, 2010) を引き起こすABCG5の両方のコピーにおける非連鎖の不活性化突然変異の発見によって実証されるように、ハプロタイプフェージングデータを有することによって、特定の遺伝子への疾患の関連がより容易に明らかにされる。

【0134】

ヒトは、1,000において平均1部位でヘテロ接合性である。場合によっては、ハイスループット配列決定法を使用するデータの単一レーンは、少なくとも約150,000,000のリードを生成する。さらなる場合において、個々のリードは、約100塩基対長である。本発明者らは、入力DNA断片が平均150kbpのサイズであると仮定し、断片当たり100のペアエンドリードを得る場合、セット当たり、すなわち100のリード対当たり30のヘテロ接合性部位を観察すると予想する。セット内にヘテロ接合部位を含むすべてのリード対は、同じセット内の他のすべてのリード対に対して相同性である (すなわち、分子的に結合される)。この特性は、場合によっては、リードの単数の対とは対照的に、セット内のフェージングのより大きな力を可能にする。ヒトゲノム中の約30億塩基、および1000中の1つがヘテロ接合性である場合、平均ヒトゲノム中に約300万のヘテロ接合性部位が存在する。ヘテロ接合性部位を含む約45,000,000個のリード対では、ハイスループット配列決定法の単一レーンを使用してフェージングされる各ヘテロ接合性部位の平均カバレッジは、典型的なハイスループット配列決定機を使用して、約(15倍(15X))である。したがって、二倍体ヒトゲノムは、本明細書に開

示される方法を使用して調製される試料由来の配列バリエーションに関連するハイスループット配列データの1つのレーンと、確実にかつ完全にフェージングすることができる。場合によっては、データのレーンは、DNA配列リードデータのセットである。さらなる場合において、データのレーンは、ハイスループット配列決定機の1回のランからのDNA配列リードデータのセットである。

【0135】

ヒトゲノムは染色体の2つの相同セットからなるため、個体の真の遺伝子構成を理解することは、遺伝物質の母体および父体のコピーまたはハプロタイプの描写を必要とする。個体においてハプロタイプを得ることは、いくつかの方法で有用である。例えば、ハプロタイプは、臓器移植におけるドナー-宿主適合の結果の予測において臨床的に有用である。ハプロタイプは、疾患関連性を検出するためにますます使用されている。複合ヘテロ接合性を示す遺伝子において、ハプロタイプは、2つの有害なバリエーションが同じ対立遺伝子（すなわち、遺伝学用語を使用するなら、「シス」）上に位置するか、または2つの異なる対立遺伝子（「トランス」）上に位置するかに関する情報を提供し、これらのバリエーションの遺伝が有害であるかどうかの予測に大きく影響し、個体が機能的対立遺伝子および2つの有害なバリエーション位置を有する単一の非機能的対立遺伝子を有するかどうか、または、その個体がそれぞれ異なる欠陥を有する2つの非機能的対立遺伝子を有するかどうかについて、結論に影響を与える。個体群由来のハプロタイプは、疫学者および人体学者の両方の関心対象である集団構造に関する、かつ、人類の進化の歴史に有益な情報を提供している。加えて、遺伝子発現における広範な対立遺伝子の不均衡が報告されており、対立遺伝子フェーズ間の遺伝的または後成的な差異が発現の定量的差異に寄与する可能性があることを示唆している。ハプロタイプ構造の理解は、対立遺伝子不均衡に寄与するバリエーションの機構を描写するであろう。

【0136】

特定の実施形態では、本明細書に開示される方法は、ロングレンジの結合およびフェージングのために必要なゲノムの遠位領域間の会合を固定および捕捉するためのインビトロ技術を含む。場合によっては、方法は、非常にゲノム的に遠位のリード対を送達するために、1つ以上のリードセットを構築および配列決定する工程を含む。さらなる場合において、各リードセットは、共通のバーコードによって標識された2つ以上のリードを含み、これは、共通のポリヌクレオチドからの2つ以上の配列セグメントを表し得る。場合によっては、相互作用は主として単一のポリヌクレオチド内のランダムな会合から生じる。場合によっては、ポリヌクレオチドにおいて互いに近接する配列セグメントがより頻繁にかつより高い確率で相互作用するが、分子の遠位部分間の相互作用はより低頻度であるため、配列セグメント間のゲノム距離が推測される。したがって、2つの遺伝子座を結合する対の数と入力DNA上でのそれらの近接性との間には系統的な関係がある。

【0137】

いくつかの態様では、本開示は、極めて高いフェージング精度を達成するためのデータを生成する方法および組成物を提供する。以前の方法と比較して、本明細書に記載される方法は、より高い割合のバリエーションをフェージングすることができる。場合によっては、高レベルの精度を維持しながらフェージングが達成される。さらなる場合において、このフェージング情報は、例えば、約200kbp、約300kbp、約400kbp、約500kbp、約600kbp、約700kbp、約800kbp、約900kbp、約1Mbp、約2Mbp、約3Mbp、約4Mbp、約5Mbp、または約10Mbpを超えて、あるいは、最大で染色体の全長を含んで、約10Mbpを超えるロングレンジに拡張される。いくつかの実施形態では、ヒト試料のヘテロ接合性SNPの90%超が、約2.5億未満のリードを使用して、例えば、Illumina HiSeqデータの1レーンのみを使用することによって、99%を超える精度でフェージングされる。他の場合において、ヒト試料についてのヘテロ接合性SNPの約40%、50%、60%、70%、80%、90%、95%、または99%超が、約2.5億未満または約5億未満のリードを使用して、例えば、Illumina HiSeqデータの1または2レーンのみを使用

することによって、約70%、80%、90%、95%、または99%を超える精度で、フェージングされる。場合によっては、ヒト試料についてのヘテロ接合性SNPの95%または99%超が、約2.5億または約5億未満のリードを用いて、約95%または99%を超える精度でフェージングされる。さらなる場合において、追加のバリエーションは、リード長を約200bp、250bp、300bp、350bp、400bp、450bp、500bp、600bp、800bp、1000bp、1500bp、2kbp、3kbp、4kbp、5kbp、10kbp、20kbp、50kbp、または100kbpに増加させることによって捕捉される。

【0138】

本開示の組成物および方法は、遺伝子発現分析に使用され得る。本明細書に記載される方法は、ヌクレオチド配列を区別する。標的ヌクレオチド配列間の差は、例えば、単一の核酸塩基の差、核酸欠失、核酸挿入、または再編成であり得る。1より多くの塩基を含むそのような配列の差も検出され得る。本開示のプロセスは、感染症、遺伝子疾患、および癌を検出することができる。さらに、上記プロセスは環境モニタリング、法医学、および食品科学においても有用である。核酸に対して実施され得る遺伝子分析の例としては、例えば、SNP検出、STR検出、RNA発現分析、プロモーターメチル化、遺伝子発現、ウイルス検出、ウイルスのサブタイプ分類、および薬物耐性が挙げられる。

【0139】

本方法は、罹患した細胞型が試料に存在するかどうか、疾患の段階、患者の予後、特定の処置に回答する患者の能力、または患者にとって最良の処置を判定するために、患者から得られたまたは患者に由来する有生分子試料の分析に適用され得る。本方法は、特定の疾患のためのバイオマーカーを同定するためにも適用され得る。

【0140】

いくつかの実施形態において、本明細書に記載される方法は疾病の診断に使用される。本明細書で使用されるように、「診断する」もしくは疾病の「診断」という用語は、疾病を予測もしくは診断すること、疾病の素因を判定すること、疾病の処置をモニタリングすること、疾患の治療反応、もしくは疾病の予後、疾病の進行、または疾病の特定の処置に対する反応を診断することを含む。例えば、血液試料は、試料中の疾患または悪性細胞型のマーカーの存在および／または量を判定するために、本明細書に記載される方法のうちいずれかに従ってアッセイすることができ、それによって、疾患または癌を診断またはステージ分類することができる。

【0141】

いくつかの実施形態において、本明細書に記載される方法および組成物は、疾病の診断および予後のために使用される。

【0142】

多数の免疫学的、増殖性、および悪性の疾患と障害は、本明細書に記載される方法に特に適している。免疫疾患および障害は、アレルギー性疾患および障害、免疫機能の障害、ならびに自己免疫疾患および疾病を含む。アレルギー性疾患および障害は、限定されないが、アレルギー性鼻炎、アレルギー性結膜炎、アレルギー性喘息、アトピー性湿疹、アトピー性皮膚炎、および食物アレルギーを含む。免疫不全症は、限定されないが、重症複合免疫不全症(SCID)、好酸球増加症候群、慢性肉芽腫症、白血球接着不全症IおよびII、高IgE症候群、チェディアック・東、好中球増加症、好中球減少症、無形成症、無ガンマグロブリン血症、高IgM症候群、ディジョージ／軟口蓋帆・心臓・顔症候群、およびインターフェロン・ガンマーTH1経路欠損症を含む。自己免疫性および免疫調節異常の障害は、限定されないが、関節リウマチ、糖尿病、全身性エリトマトーデス、グレーブス病、グレーブス眼症、クローン病、多発性硬化症、乾癬、全身性硬化症、甲状腺腫およびリンパ腫性甲状腺腫(橋本甲状腺炎、リンパ節様甲状腺腫)、円形脱毛症、自己免疫性心筋炎、硬化性苔癬、自己免疫性ブドウ膜炎、アジソン病、萎縮性胃炎、重症筋無力症、特発性血小板減少性紫斑病、溶血性貧血、原発性胆汁性肝硬変、ウェゲナー肉芽腫症、結節性多発動脈炎、および炎症性腸疾患、同種異系移植片拒絶反応、ならびに感染性細

菌または環境抗原に対するアレルギー反応による組織破壊を含む。

【0143】

本開示の方法により評価され得る増殖性疾患および障害は、限定されないが、新生児の血管腫症、二次性進行型多発性硬化症、慢性進行性骨髄変性疾患、神経線維腫症、神経節神経腫症、ケロイド形成、骨のパジエット病、（例えば、乳房または子宮の）線維嚢胞症、サルコイドーシス、ペロニーおよびデュピュイトランの繊維症、硬変、アテローム性動脈硬化、および血管再狭窄を含む。

【0144】

本開示の方法により評価され得る悪性疾患および障害は、血液悪性腫瘍および固形腫瘍の両方を含む。

【0145】

血液悪性腫瘍は特に、血液由来の細胞の変化を含むため、試料が血液試料である時に本開示の方法に適している。そのような悪性腫瘍は、非ホジキンリンパ腫、ホジキンリンパ腫、非B細胞リンパ腫、および他のリンパ腫、急性または慢性の白血病、多血球血症、血小板血症、多発性骨髄腫、骨髄異形成障害、骨髄増殖性障害、骨髄線維症（myelofibroses）、異型免疫リンパ球増殖、および形質細胞障害を含む。

【0146】

本開示の方法により評価され得る形質細胞疾患は、多発性骨髄腫、アミロイドーシス、およびヴァルデンストレームマクログロブリン血症を含む。

【0147】

固形腫瘍の例としては、限定されないが、結腸癌、乳癌、肺癌、前立腺癌、脳腫瘍、中枢神経系腫瘍、膀胱腫瘍、黒色腫、肝臓癌、骨肉腫、および他の骨癌、睾丸および卵巣の癌腫、頭頸部腫瘍、および子宮頸部の新生物が挙げられる。

【0148】

遺伝子疾患も本開示のプロセスにより検出することができる。これは、染色体および遺伝子の異常、または遺伝子疾患に関する出生前または出生後のスクリーニングにより実行され得る。検出可能な遺伝子疾患の例は、21ヒドロキシラーゼ欠損、嚢胞性繊維症、脆弱X症候群、ターナー症候群、デュシェンヌ型筋ジストロフィー、ダウン症候群または他のトリソミー、心疾患、単一の遺伝子疾患、HLA分類、フェニルケトン尿症、鎌状赤血球貧血、テイ・ザックス病、サラセミア、クライنفエルター症候群、ハンチントン病、自己免疫疾患、リポドーシス、肥満欠損症（obesity defect）、血友病、先天性代謝異常症、および糖尿病を含む。

【0149】

本明細書に記載される方法は、試料中の細菌またはウイルスそれぞれのマーカーの存在および／または量を判定することにより、病原体感染、例えば、細胞内細菌およびウイルスによる感染を診断するために使用され得る。

【0150】

多種多様な感染症が本開示のプロセスにより検出することができる。感染症は、細菌、ウイルス、寄生生物、および真菌の感染因子により引き起こされ得る。薬物に対する様々な感染因子の抵抗性も、本開示を使用して判定され得る。

【0151】

本開示により検出され得る細菌感染因子は、エシェリキア・コリ、サルモネラ、シゲラ、クレブシエラ、シュドモナス、リステリア・モノサイトゲネス、マイコバクテリウム・ツベルクローシス、マイコバクテリウム・アビウム・イントラセルラーレ、エルシニア、フランシセラ、パスツレラ、ブルセラ、クロストリジウム、ボルデテラ・ペルツシス、バクテロイデス、スタフィロコッカス・アウレウス、ストレプトコッカス・ニューモニエ、B溶血性連鎖球菌（B-Hemolytic strep.）、コリネバクテリア、レジオネラ、ミコプラズマ、ウレアプラスマ、クラミジア、ナイセリア・ゴノレエ（Neisseria gonorrhoeae）、ナイセリア・メニンギティディス、ヘモフィルス・インフルエンザ、エンテロコッカス・フェカリス、プロテウス・ブルガリス、ブ

10

20

30

40

50

ロテウス・ミラビリス、ヘリコバクターピロリ、トレポネーマ・パラジウム、ボレリア・ブルグドルフェリ、ボレリア・リカレンチス、リケッチア病原体、ノカルディア、および放線菌（*Actinomycetes*）を含む。

【0152】

本開示により検出され得る真菌感染因子は、クリプトコッカス・ネオフォルマンズ、ブラストミセス・デルマトフィティディス、ヒストプラスマ・カプスラーツム、コクシジオイデス・イミチス、パラコクシジオイデス・ブラジリエンシス、カンジダ・アルビカンズ、アスペルギルス・フミガーツス（*Aspergillus fumigatus*）、藻菌類（リゾープス）、スポロトリックス・シェンキー、クロモミコーシス、およびマズラミコーシスを含む。

10

【0153】

本開示により検出されるウイルス感染因子は、ヒト免疫不全ウイルス、ヒトT細胞リンパ性細胞栄養性ウイルス（*human T-cell lymphocytotropic virus*）、肝炎ウイルス（例えば、B型肝炎ウイルスおよびC型肝炎ウイルス）、エプスタイン・バーウイルス、サイトメガロウイルス、ヒトパピローマウイルス、オルソミクソウイルス、パラミクソウイルス、アデノウイルス、コロナウイルス、ラブドウイルス、ポリオウイルス、トーガウイルス、ブンヤウイルス、アレナウイルス、風疹ウイルス、およびレオウイルスを含む。

【0154】

本開示により検出され得る寄生生物因子は、熱帯熱マラリア原虫、四日熱マラリア原虫、三日熱マラリア原虫、卵型マラリア原虫、回旋糸状虫（*Onchocerca volvulus*）、リーシュマニア、トリパノゾーマ種、住血吸虫種、エントアメーバーヒストリチカ、クリプトスポリジウム（*Cryptosporidium*）、ジアルジア種、トリコモナス（*Trichomonas*）種、大腸バランチジウム（*Balantidium Coli*）、バンクロフト糸状虫、トキソプラズマ種、蟯虫、回虫、鞭虫、メジナ虫（*Dracunculus medinensis*）、吸虫、広節裂頭条虫、テニア種、ニューモシスチス・カリニ、およびアメリカ鉤虫（*Necator americanus*）を含む。

20

【0155】

本開示は、感染因子による薬物抵抗性の検出にも有用である。例えば、バイコマイシン耐性エンテロコッカス・フェシウム、メチシリン耐性スタフィロコッカス・アウレウス、ペニシリン耐性ストレプトコッカス・ニューモニエ、多剤耐性マイコバクテリウム・ツベルクローシス、および耐AZT性ヒト免疫不全ウイルスはすべて、本開示で同定され得る。

30

【0156】

したがって、本開示の組成物および方法を使用して検出された標的分子は、患者のマーカー（癌マーカーなど）、または細菌もしくはウイルスマーカーなどの異物による感染のマーカーのいずれかであり得る。

【0157】

本開示の組成物および方法は、標的分子を同定および／または定量化するために使用され得、標的分子の存在量は、生物学的状態または疾患条件、例えば、疾患状態の結果としてアップレギュレートまたはダウンレギュレートされる血液マーカーを示す。

40

【0158】

いくつかの実施形態において、本開示の方法および組成物は、サイトカイン発現のために使用することができる。本明細書に記載される方法の低感受性は、例えば、癌などの疾患の状態、診断、または予後のバイオマーカーとしてのサイトカインの早期発見、および亜臨床的な状態の同定に有用である。

【0159】

標的ポリヌクレオチドが由来する様々な試料は、同じ個体の複数の試料、異なる個体の試料、あるいはそれらの組み合わせを含み得る。いくつかの実施形態において、試料は、

50

1人の個体の複数のポリヌクレオチドを含む。いくつかの実施形態において、試料は、2人以上の個体の複数のポリヌクレオチドを含む。個体は、標的ポリヌクレオチドが由来し得る任意の生物またはその一部であり、その非限定的な例としては、植物、動物、真菌、原生生物、モネラ界、ウイルス、ミトコンドリア、および葉緑体が挙げられる。試料ポリヌクレオチドは、細胞試料、組織試料、または、例えば、培養細胞株、生検、血液試料、もしくは細胞を含有している流体試料を含む、それらに由来する臓器試料など、対象から単離することができるものである。対象は、限定されないが、ウシ、ブタ、マウス、ラット、ニワトリ、ネコ、イヌなどの動物を含む動物であってもよく、通常はヒトなどの哺乳動物である。試料は化学合成などにより人工的に得られることもできる。いくつかの実施形態において、試料はDNAを含む。いくつかの実施形態において、試料はゲノムDNAを含む。いくつかの実施形態において、試料は、ミトコンドリアDNA、葉緑体DNA、プラスミドDNA、バクテリア人工染色体、酵母人工染色体、オリゴヌクレオチドタグ、またはそれらの組み合わせを含む。いくつかの実施形態において、試料は、限定されないが、ポリメラーゼ連鎖反応（PCR）、逆転写、およびそれらの組み合わせを含む、プライマーとDNAポリメラーゼとの適切な組み合わせを使用したプライマー伸長反応により生成されたDNAを含む。プライマー伸長反応のための鋳型がRNAである場合、逆転写の産物は相補的DNA（cDNA）と呼ばれる。プライマー伸長反応に有用なプライマーは、1つ以上の標的に特異的な配列、ランダム配列、部分的なランダム配列、およびそれらの組み合わせを含み得る。プライマー伸長反応に適した反応条件は当該技術分野で知られている。一般に、試料のポリヌクレオチドは、試料中に存在する任意のポリヌクレオチドを含み、これは標的ポリヌクレオチドを含むこともあれば、含まないこともある。

10

20

【0160】

いくつかの実施形態において、核酸鋳型分子（例えば、DNAまたはRNA）は、タンパク質、脂質、および非鋳型核酸などの、様々な他の成分を含有する生体試料から単離される。核酸鋳型分子は任意の細胞材料から得られ、動物、植物、細菌、真菌類、または他の細胞生物から得られ得る。本開示での使用のための生体試料は、ウイルスの粒子または調製物を含む。核酸鋳型分子は、生物から直接、または、生物から得た生体試料、例えば、血液、尿、脳脊髄液、精液、唾液、痰、便、および組織から得ることができる。任意の組織または体液の標本が、本開示で使用される核酸の供給源として使用されてもよい。核酸鋳型分子は、初代細胞培養物または細胞株などの培養細胞からも単離され得る。鋳型核酸が得られる細胞または組織は、ウイルスまたは他の細胞内病原体に感染され得る。試料は、生体標本、cDNAライブラリー、ウイルスDNA、またはゲノムDNAから抽出された総体的なRNAでもあり得る。試料は、細胞構造がない起源から単離されたDNA、例えば、冷凍装置から増幅／単離されたDNAでもあり得る。

30

【0161】

核酸の抽出と精製のための方法は公知である。例えば、核酸は、フェノール、フェノール／クロロホルム／イソアミルアルコール、あるいはTRIzolおよびTriReagentを含む同様の製剤での有機抽出により精製され得る。抽出技術の他の限定されない例は、（1）自動核酸抽出器、例えば、Applied Biosystems（Foster City, Calif.）から入手可能なModel 341 DNA Extractorの使用を伴うまたは伴わない、例えば、フェノール／クロロホルムの有機試薬（Ausubel et al., 1993）を使用する、有機抽出とその後のエタノール沈殿、（2）固定相吸着法（米国特許第5,234,809号、Walsh et al., 1991）、および（3）典型的に「塩析」方法と呼ばれる沈澱法などの、塩で誘導された核酸沈澱法（Miller et al., (1988)）を含む。核酸の単離および／または精製の別の例は、核酸が特異的または非特異的に結合することができる磁性粒子の使用と、その後の、磁石を使用したビーズの単離と、核酸の洗浄と、ビーズからの核酸の溶出を含む（例えば、米国特許第5,705,628号を参照）。いくつかの実施形態において、上記の単離方法は、試料から不要なタンパク質を取り除くのに役立つ酵素消化工程、例えば、プロテイナーゼKまたは他の同様のプロテアーゼによる消化によ

40

50

って始められてもよい。例えば、米国特許第7,001,724号を参照。必要に応じて、RNase阻害剤を溶解緩衝液に添加することができる。特定の細胞または試料型について、前記プロトコルにタンパク質変性／消化の工程を加えることが望ましい場合もある。精製方法は、DNA、RNA、またはその両方を単離することを目的とし得る。抽出手順の間またはその後、DNAとRNAの両方が一緒に単離されると、さらなる工程を利用して、一方または両方を他とは別々に精製することができる。例えば、サイズ、配列、または他の物理的もしくは化学的な特性による精製により、抽出した核酸の細画分を生成することもできる。最初の核酸単離工程に加えて、過剰なまたは不要な試薬、反応物、または生成物を除去するなどのために、本開示の方法における工程の後に、核酸の精製を実施することができる。

10

【0162】

核酸の鑄型分子は、2003年10月9日に公開された米国特許出願公開第US2002/0190663号A1に記載の通り得ることができる。通常、核酸は、Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281 (1982)に記載されるものなどの様々な技術によって生体試料から抽出することができる。場合によっては、核酸は、最初に生体試料から抽出され、次いでインビトロでクロスリンクされ得る。場合によっては、天然の会合タンパク質（例えば、ヒストン）を、核酸からさらに除去することができる。

20

【0163】

他の実施形態では、本開示は、例えば、組織、細胞培養物、体液、動物組織、植物、細菌、真菌、ウイルスなどから単離されたDNAを含む任意の高分子量二本鎖DNAに容易に適用することができる。

【0164】

サイズ選択を含むHi-C方法

【0165】

少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料をDNaseと接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程と、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程と、を含む方法が本明細書で提供される。場合によっては、複数の選択されたセグメントは約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100bp～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpである。

30

40

【0166】

本明細書に提供されるサイズ選択工程を含む方法の別の態様では、方法は、サイズ選択工程の前に、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含む。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含む。場合によっては、サイズ選択されたライブラリーは、約350bp～約1000bpのサイズである。場合によっては、サイズ選択ライブラリーは、約100bp～約2500bpのサイズ、例えば、約100b

50

p ～約 3 5 0 b p、約 3 5 0 b p ～約 5 0 0 b p、約 5 0 0 b p ～約 1 0 0 0 b p、約 1 0 0 0 b p ～約 1 5 0 0 b p および約 2 0 0 0 b p、約 2 0 0 0 b p ～約 2 5 0 0 b p、約 3 5 0 b p ～約 1 0 0 0 b p、約 3 5 0 b p ～約 1 5 0 0 b p、約 3 5 0 b p ～約 2 0 0 0 b p、約 3 5 0 b p ～約 2 5 0 0 b p、約 5 0 0 b p ～約 1 5 0 0 b p、約 5 0 0 b p ～約 2 0 0 0 b p、約 5 0 0 b p ～約 3 5 0 0 b p、約 1 0 0 0 b p ～約 1 5 0 0 b p、約 1 0 0 0 b p ～約 2 0 0 0 b p、約 1 0 0 0 b p ～約 2 5 0 0 b p、約 1 5 0 0 b p ～約 2 0 0 0 b p、約 1 5 0 0 b p ～約 2 5 0 0 b p、または約 2 0 0 0 b p ～約 2 5 0 0 b p である。

【 0 1 6 7 】

本明細書に提供されるサイズ選択工程を含む方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、他の適切な方法、またはそれらの組み合わせを用いて行うことができる。

【 0 1 6 8 】

別の態様では、本明細書で提供されるサイズ選択工程を含む方法は、Q C 値を得るために複数の選択されたセグメントを分析する工程をさらに含むことができる。場合によっては、Q C 値は、クロマチン消化効率 (C D E) およびクロマチン消化指数 (G D I) から選択される。C D E は、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、C D E は、サイズ選択前の 1 0 0 ～ 2 5 0 0 b p のサイズのセグメントの割合として計算される。場合によっては、試料は、C D E 値が少なくとも 6 5 % であるときに、さらなる分析のために選択される。場合によっては、試料は、C D E 値が少なくとも約 5 0 %、少なくとも約 5 5 %、少なくとも約 6 0 %、少なくとも約 6 5 %、少なくとも約 7 0 %、少なくとも約 7 5 %、少なくとも約 8 0 %、少なくとも約 8 5 %、少なくとも約 9 0 %、または少なくとも約 9 5 % である場合に、さらなる分析のために選択される。C D I は、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算される。例えば、C D I は、6 0 0 ～ 2 5 0 0 b p のサイズを有する断片対 1 0 0 ～ 6 0 0 b p のサイズを有する断片の比の対数として計算され得る。場合によっては、試料は、C D I 値が - 1 . 5 より大きくかつ 1 未満である場合に、さらなる分析のために選択される。場合によっては、C D I 値が約 - 2 より大きく約 1 . 5 未満、約 - 1 . 9 より大きく約 1 . 5 未満、約 - 1 . 8 より大きく約 1 . 5 未満、約 - 1 . 7 より大きく約 1 . 5 未満、約 - 1 . 6 より大きく約 1 . 5 未満、約 - 1 . 5 より大きく約 1 . 5 未満、約 - 1 . 4 より大きく約 1 . 5 未満、約 - 1 . 3 より大きく約 1 . 5 未満、約 - 1 . 2 より大きく約 1 . 5 未満、約 - 1 . 1 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 5 未満、約 - 1 より大きく約 1 . 5 未満、約 - 0 . 9 より大きく約 1 . 5 未満、約 - 0 . 8 より大きく約 1 . 5 未満、約 - 0 . 7 より大きく約 1 . 5 未満、約 - 0 . 6 より大きく約 1 . 5 未満、約 - 0 . 5 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 4 未満、約 - 2 より大きく約 1 . 3 未満、約 - 2 より大きく約 1 . 2 未満、約 - 2 より大きく約 1 . 1 未満、約 - 2 より大きく約 1 未満、約 - 2 より大きく約 0 . 9 未満、約 - 2 より大きく約 0 . 8 未満、約 - 2 より大きく約 0 . 7 未満、約 - 2 より大きく約 0 . 6 未満、または約 - 2 より大きく約 0 . 5 未満である場合に、試料はさらなる分析のために選択される。

【 0 1 6 9 】

別の態様において、本明細書におけるサイズ選択工程を含む方法において使用される安定化された生体試料は、安定化剤で処理された生体材料を含む。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。代替的に、安定化された生体試料は安定化された無傷細胞を含む。代替的に、安定化された生体試料は安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核試料を D N a s e に接触させる工程は、無傷細胞または無傷の核の溶解の前に行われる。場合によっては、細胞および／または核は、接合部において複数のセグメントの第 1 のセグメントおよび第 2 のセグメントを結合させる前に溶解される。

【 0 1 7 0 】

別の態様において、本明細書におけるサイズ選択工程を含む方法は、少数の細胞または少量の核酸を含有する少量の試料において行われる。例えば、場合によっては、安定化された生体試料は3,000,000個未満の細胞を含む。場合によっては、安定化された生体試料は2,000,000個未満の細胞を含む。場合によっては、安定化された生体試料は1,000,000個未満の細胞を含む。場合によっては、安定化された生体試料は500,000個未満の細胞を含む。場合によっては、安定化された生体試料は400,000個未満の細胞を含む。場合によっては、安定化された生体試料は300,000個未満の細胞を含む。場合によっては、安定化された生体試料は200,000個未満の細胞を含む。場合によっては、安定化された生体試料は100,000個未満の細胞を含む。場合によっては、安定化された生体試料は50,000個未満の細胞を含む。場合によっては、安定化された生体試料は40,000個未満の細胞を含む。場合によっては、安定化された生体試料は30,000個未満の細胞を含む。場合によっては、安定化された生体試料は20,000個未満の細胞を含む。場合によっては、安定化された生体試料は10,000個の細胞を含むことができる。場合によっては、安定化された生体試料は10 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は9 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は8 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は7 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は6 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は5 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は4 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は3 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は2 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は1 μ g 未満のDNAを含む。場合によっては、安定化された生体試料は0.5 μ g 未満のDNAを含む。

10

20

30

40

50

【0171】

別の態様では、本明細書のサイズ選択工程を含む方法は、個々の細胞または単一の細胞に対して行うことができる。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【0172】

さらなる態様では、本明細書のサイズ選択工程を含む方法で使用される安定化された生体試料は、DNAの断片を作製するためにDNaseなどのヌクレアーゼで処理される。場合によっては、DNaseは非配列特異的である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性である。場合によっては、DNaseは二本鎖DNAに特異的である。場合によっては、DNaseは二本鎖DNAを優先的に切断する。場合によっては、DNaseは一本鎖DNAに特異的である。場合によっては、DNaseは一本鎖DNAを優先的に切断する。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseは小球菌ヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼのうちの1つ以上から選択される。場合によっては、DNaseは、免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。免疫グロブリン結合タンパク質は、例えば、プロテインA、プロテインG、プロテインA/G、またはプロテインLであり得る。いくつかの実施形態において、DNaseは、2つ以上の免疫グロブリン結合タンパク質および／またはその断片を含む融合タンパク質に結合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

【0173】

さらなる態様において、サイズ選択工程を含む方法において使用するための本明細書に提供される安定化された生体試料は、1つ以上のクロスリンク剤で処理される。場合によっては、クロスリンク剤は化学的固定剤である。場合によっては、化学的固定剤は、約2.3〜2.7オングストローム(Å)のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または20 Åのスペーサー長を有することができる。化学的固定剤は、約16.1 Åの長さを有するスペーサーアームを有するエチレングリコールビス(スクシンイミジルスクシネート)(EGS)を含むことができる。化学的固定剤は、約7.7 Åの長さのスペーサーアームを有するジスクシンイミジルグルタレート(DSG)を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよびEGS、ホルムアルデヒドおよびDSG、またはホルムアルデヒド、EGS、およびDSGを含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用され、他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな(例えば、>1 kb)リード対分離距離を有するリード対の割合を増加させることができる。例えば、図7は、得られたライブラリー(消化されたDNaseおよびMNaseの両方)をホルムアルデヒド単独でクロスリンクしたものと、ホルムアルデヒドとDSGまたはEGSでクロスリンクしたものとを比較を示す。DSGは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。DSGは膜透過性であり、細胞内クロスリンクを可能にする。DSGは、いくつかの用途においてジスクシンイミジルスベレート(DSS)と比較して、クロスリンク効率を高めることができる。EGSは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。EGSは膜透過性であり、細胞内クロスリンクを可能にする。EGSクロスリンクは、例えば、pH8.5で3〜6時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGSとの可逆的クロスリンク後にその活性の60%を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、クロスリンク剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファミド(isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン(ciccy cloplatin)、エプタプラチン(eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド(mitozolomide)、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンである。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。

10

20

30

40

50

【0174】

さらなる態様では、本明細書で提供されるサイズ選択工程を含む方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0175】

さらなる態様では、本明細書で提供されるサイズ選択工程を含む方法は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端をfill-inすること、および平滑末端をライゲーションすることを含む。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを架

橋オリゴヌクレオチドに接触させることを含む。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約15～約18ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約25、約30、約35、約40、約45、または約50ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含み得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含むことができる。いくつかの実施形態では、架橋オリゴヌクレオチドは、互いに結合された複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合され(coupledまたはlinked)得る。場合によっては、結合された架橋オリゴヌクレオチドは、抗体が結合される試料核酸中の位置に送達され得る。

10

20

30

40

【0176】

分割およびプーリングアプローチを用いて、固有のバーコードを有する架橋オリゴヌクレオチドを生成することができる。試料の集団は、複数の群に分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドバーコードが群間で異なるが、群内では同じであるように試料に結合させることができ、試料の群を再度プールすることができ、このプロセスを複数回繰り返すことができる。このプロセスを反復すると、最終的には、集団中の各試料が固有の一連の架橋オリゴヌクレオチドバーコードを有することになり、単一試料(例えば、単細胞、単核、単染色体)の分析が可能になる。1つの例示的な例では、ビーズの固体支持体に結合したクロスリンクした消化核の試料を8本のチューブに分け、各々が、ライゲーションされる二本鎖DNA(dsDNA)アダプターを含む第1のアダプター群(第1の反復)の8つの固有のメンバーのうちの1つを含有する。8つのアダプターの各々は、核中のクロスリンクしたクロマチン凝集体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有することができるが、そうでなければ固有のdsDNA配列を有する。第1のアダプター群をライゲーションした後、核をプールして戻し、洗浄してライゲーション反応成分を除去することができる。分配、ライゲーション、およびプールのスキームは、さらに2回繰り返すことができる(2回の反復)。各アダプター群からのメンバーのライゲーション後に、クロスリンクしたクロマチン凝集体を複数のバーコードに連続して結合させることができる。場合によっては、複数のアダプター群の複数のメンバーの連続ライゲーション(反復)は、バーコードの組み合わせをもたらす。利用可能なバーコードの組み合わせの数は、反復当たりの群の数、および使用されるバーコードオリゴヌクレオチドの総数に依存する。例えば、それぞれ8つのメンバーを含む3つの反復は、83の可能な組み合わせを有することができる。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加または減少させることによって、および/または、反復の数を増加または減少させることによって調整することができる。1を超えるアダプター群が使用される場合、反復アダプター結合のために、分配、結合、およびプールのスキームを使用することができる。場合によっては、分配、結合、およびプールのスキームは、少なくとも3、4、5、6、7、8、9、または10回さらに繰り返され得る。場合によっては、最後のアダプター群のメンバーは、例えば、PCR増幅による配列決定ライブラリー調製中に、アダプターに結合したDNAのその後の濃縮のための配列を含む。

【0177】

さらなる態様において、本明細書におけるサイズ選択工程を含む方法は、せん断工程を含まない(例えば、核酸はせん断されない)。

50

10

20

30

40

50

【 0 1 8 1 】

【 0 1 8 2 】

【 0 1 8 3 】

さらに、少なくとも１つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料をDNaseと接触させる工程と、接合部において複数のセグメントの第１のセグメントおよび第２のセグメントを結合させる工程と、QC値を決定するために複数のセグメントを分析する工程と、を含む方法が本明細書で提供される。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指数(GDI)から選択される。CDEは、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、CDEは、サイズ選択前の100~2500bpのサイズのセグメントの割合として計算される。場合によっては、試料は、CDE値が少なくとも65%であるときに、さらなる分析のために選択される。場合によっては、試料は、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%である場合に、さらなる分析のために選択される。CDIは、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算される。例えば、CDIは、600~2500bpのサイズを有する断片対100~600bpのサイズを有する断片の比の対数として計算され得る。場合によっては、試料は、CDI値が-1.5より大きくかつ1未満である場合に、さらなる分析のために選択される。場合によっては、CDI値が約-2より大きく約1.5未満、約-1.9より大きく約1.5未満、約-1.8より大きく約1.5未満、約-1.7より大きく約1.5未満、約-1.6より大きく約1.5未満、約-1.5より大きく約1.5未満、約-1.4より大きく約1.5未満、約-1.3より大きく約1.5未満、約-1.2より大きく約1.5未満、約-1.1より大きく約1.

5 未満、約 - 2 より大きく約 1.5 未満、約 - 1 より大きく約 1.5 未満、約 - 0.9 より大きく約 1.5 未満、約 - 0.8 より大きく約 1.5 未満、約 - 0.7 より大きく約 1.5 未満、約 - 0.6 より大きく約 1.5 未満、約 - 0.5 より大きく約 1.5 未満、約 - 2 より大きく約 1.4 未満、約 - 2 より大きく約 1.3 未満、約 - 2 より大きく約 1.2 未満、約 - 2 より大きく約 1.1 未満、約 - 2 より大きく約 1 未満、約 - 2 より大きく約 0.9 未満、約 - 2 より大きく約 0.8 未満、約 - 2 より大きく約 0.7 未満、約 - 2 より大きく約 0.6 未満、または約 - 2 より大きく約 0.5 未満である場合に、試料はさらなる分析のために選択される。

【0184】

他の態様では、本明細書の QC 決定工程を含む方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程を含み得る。場合によっては、複数の選択されたセグメントは約 145 ~ 約 600 bp である。場合によっては、複数の選択されたセグメントは約 100 ~ 約 2500 bp である。場合によっては、複数の選択されたセグメントは約 100 ~ 約 600 bp である。場合によっては、複数の選択されたセグメントは約 600 ~ 約 2500 bp である。場合によっては、複数の選択されたセグメントは、約 100 bp ~ 約 600 bp、約 100 bp ~ 約 700 bp、約 100 bp ~ 約 800 bp、約 100 bp ~ 約 900 bp、約 100 bp ~ 約 1000 bp、約 100 bp ~ 約 1100 bp、約 100 bp ~ 約 1200 bp、約 100 bp ~ 約 1300 bp、約 100 bp ~ 約 1400 bp、約 100 bp ~ 約 1500 bp、約 100 bp ~ 約 1600 bp、約 100 bp ~ 約 1700 bp、約 100 bp ~ 約 1800 bp、約 100 bp ~ 約 1900 bp、約 100 bp ~ 約 2000 bp、約 100 bp ~ 約 2100 bp、約 100 bp ~ 約 2200 bp、約 100 bp ~ 約 2300 bp、約 100 bp ~ 約 2400 bp、または約 100 bp ~ 約 2500 bp である。

【0185】

本明細書に提供される QC 決定工程を含む方法の別の態様では、方法は、サイズ選択工程の前に、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含むことができる。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含む。場合によっては、サイズ選択されたライブラリーは、約 350 bp ~ 約 1000 bp のサイズである。場合によっては、サイズ選択ライブラリーは、約 100 bp ~ 約 2500 bp のサイズ、例えば、約 100 bp ~ 約 350 bp、約 350 bp ~ 約 500 bp、約 500 bp ~ 約 1000 bp、約 1000 bp ~ 約 1500 bp および約 2000 bp、約 2000 bp ~ 約 2500 bp、約 350 bp ~ 約 1000 bp、約 350 bp ~ 約 1500 bp、約 350 bp ~ 約 2000 bp、約 350 bp ~ 約 2500 bp、約 500 bp ~ 約 1500 bp、約 500 bp ~ 約 2000 bp、約 500 bp ~ 約 3500 bp、約 1000 bp ~ 約 1500 bp、約 1000 bp ~ 約 2000 bp、約 1000 bp ~ 約 2500 bp、約 1500 bp ~ 約 2000 bp、約 1500 bp ~ 約 2500 bp、または約 2000 bp ~ 約 2500 bp である。

【0186】

本明細書の QC 決定工程を含む方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせを用いて行うことができる。他の適切なサイズ選択の方法も本開示の範囲内である。

【0187】

別の態様において、本明細書の QC 決定工程を含む方法において使用される安定化された生体試料は、安定化剤で処理されている生体材料を含む。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。代替的に、安定化された生体試料は安定化された無傷細胞を含む。代替的に、安定化された生体試料は安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核試料を DNase に接触させる工程は、無傷細胞または無傷の核の溶解の前に行われる。場合によっては、細胞および／または核は、接合部において複数のセグメントの第 1 のセグメントおよび第 2 のセグメン

トを結合させる前に溶解される。

【 0 1 8 8 】

別の態様において、本明細書のQC決定工程を含む方法は、少数の細胞または少量の核酸を含有する少量の試料において行われる。場合によっては、安定化された生体試料は3,000,000未満の細胞を含む。場合によっては、安定化された生体試料は2,000,000個未満の細胞を含む。場合によっては、安定化された生体試料は1,000,000未満の細胞を含む。場合によっては、安定化された生体試料は500,000個未満の細胞を含む。場合によっては、安定化された生体試料は400,000個未満の細胞を含む。場合によっては、安定化された生体試料は300,000個未満の細胞を含む。場合によっては、安定化された生体試料は200,000個未満の細胞を含む。場合によっては、安定化された生体試料は100,000未満の細胞を含む。場合によっては、安定化された生体試料は50,000個未満の細胞を含む。場合によっては、安定化された生体試料は40,000個未満の細胞を含む。場合によっては、安定化された生体試料は30,000個未満の細胞を含む。場合によっては、安定化された生体試料は20,000個未満の細胞を含む。場合によっては、安定化された生体試料は10,000個未満の細胞を含む。場合によっては、安定化された生体試料は約10,000個の細胞を含む。場合によっては、安定化された生体試料は10μg未満のDNAを含む。場合によっては、安定化された生体試料は9μg未満のDNAを含む。場合によっては、安定化された生体試料は8μg未満のDNAを含む。場合によっては、安定化された生体試料は7μg未満のDNAを含む。場合によっては、安定化された生体試料は6μg未満のDNAを含む。場合によっては、安定化された生体試料は5μg未満のDNAを含む。場合によっては、安定化された生体試料は4μg未満のDNAを含む。場合によっては、安定化された生体試料は3μg未満のDNAを含む。場合によっては、安定化された生体試料は2μg未満のDNAを含む。場合によっては、安定化された生体試料は1μg未満のDNAを含む。場合によっては、安定化された生体試料は0.5μg未満のDNAを含む。

【 0 1 8 9 】

別の態様では、本明細書のQC決定工程を含む方法は、個々の細胞または単一の細胞に対して行うことができる。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【 0 1 9 0 】

さらなる態様では、本明細書のQC決定工程を含む方法で使用される安定化された生体試料は、DNAの断片を作製するためにDNaseなどのヌクレアーゼで処理される。場合によっては、DNaseは非配列特異的である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性である。場合によっては、DNaseは二本鎖DNAに特異的である。場合によっては、DNaseは二本鎖DNAを優先的に切断する。場合によっては、DNaseは一本鎖DNAに特異的である。場合によっては、DNaseは一本鎖DNAを優先的に切断する。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseは小球菌ヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼのうちの1つ以上から選択される。場合によっては、DNaseは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

【 0 1 9 1 】

さらなる態様では、本明細書のOC決定工程を含む方法で使用する安定化された生体

試料は、クロスリンク剤で処理される。場合によっては、クロスリンク剤は化学的固定剤である。場合によっては、化学的固定剤は、約 2.3 ~ 2.7 オングストローム (Å) のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約 3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または 20 Å のスペーサー長を有することができる。化学的固定剤は、約 16.1 Å の長さを有するスペーサーアームを有するエチレングリコールビス (スクシンイミジルスクシネート) (EGS) を含むことができる。化学的固定剤は、約 7.7 Å の長さのスペーサーアームを有するジスクシンイミジルグルタレート (DSG) を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよび EGS、ホルムアルデヒドおよび DSG、またはホルムアルデヒド、EGS、および、DSG を含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用される。他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな (例えば、> 1 kb) リード対分離距離を有するリード対の割合を増加させることができる。例えば、図 7 は、得られたライブラリー (消化された DNase および MNase の両方) をホルムアルデヒド単独でクロスリンクしたもの、ホルムアルデヒドと DSG または EGS でクロスリンクしたものとの比較を示す。DSG は、両方の末端に NHS エステル反応基を有し、アミノ基 (例えば、一級アミン) に対して反応性であり得る。DSG は膜透過性であり、細胞内クロスリンクを可能にする。DSG は、いくつかの用途においてジスクシンイミジルスベレート (DS) と比較して、クロスリンク効率を高めることができる。EGS は、両方の末端に NHS エステル反応基を有し、アミノ基 (例えば、一級アミン) に対して反応性であり得る。EGS は膜透過性であり、細胞内クロスリンクを可能にする。EGS クロスリンクは、例えば、pH 8.5 で 3 ~ 6 時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGS との可逆的クロスリンク後にその活性の 60 % を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、クロスリンク剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス (2-クロロエチル) エチルアミン、ビス (2-クロロエチル) メチルアミン、トリス (2-クロロエチル) アミン、イソファミド (isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン (ciccy cloplatin)、エプタプラチン (eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (mitozolomide)、テモゾロミド、マイトマイシン C、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンである。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。

10

20

30

40

【0192】

さらなる態様では、本明細書で提供される QC 決定工程を含む方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0193】

さらなる態様では、本明細書で提供される QC 決定工程を含む方法は、接合部において複数のセグメントの第 1 のセグメントおよび第 2 のセグメントを結合させる工程を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を *fish-in* すること、および平滑末端をライゲーションすることを含む。場合によっては、結合させる工程は、少なくとも第 1 のセグメントおよび第 2 のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、結合させる工程は、少な

50

くとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約15～約18ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約25、約30、約35、約40、約45、または約50ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含み得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含むことができる。いくつかの実施形態では、架橋オリゴヌクレオチドは、互いに結合された複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、プロテインG、プロテインA／G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合され(coupledまたはlinked)得る。場合によっては、結合された架橋オリゴヌクレオチドは、抗体が結合される試料核酸中の位置に送達され得る。

10

【0194】

さらなる態様では、本明細書のQC決定工程を含む方法は、せん断工程を含まない。

【0195】

本明細書のQC決定工程を含む方法のさらなる態様では、方法は、第1のリード対を生成するために、接合部の各側で少なくともいくつかの配列を得る工程を含む。例えば、方法は、第1のリード対を生成するために、接合部の各側で少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

20

【0196】

本明細書のQC決定工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とを含む。

【0197】

本明細書のQC決定工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含み得る。

30

【0198】

本明細書のQC決定工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセット中のバリエーションをフェーズに割り当てる工程とを含む。

【0199】

本明細書のQC決定工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(1)安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2)バリエーションの存在に基づいて薬物を選択すること、または、(3)安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を実施する工程とを含む。

40

【0200】

全細胞または全核消化を含むHi-C方法

少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料をDNaseと接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程であって、安定化された生体試料は無傷細胞および／または無傷の核を含む、工程とを含む方法が本明細書でさらに提供される。場合によっては、安定化された生体試料は安定化された無傷細胞を含む。代替的に、または

50

、組み合わせで、安定化された生体試料は安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核試料をDNaseに接触させる工程は、無傷細胞または無傷の核の溶解の前に行われる。場合によっては、細胞および／または核は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる前に溶解される。

【0201】

他の態様では、本明細書的全細胞または全核の消化を含む方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程を含むことができる。場合によっては、複数の選択されたセグメントは約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100bp～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpである。

【0202】

本明細書で提供される全細胞または全核の消化を含む方法の別の態様では、方法は、サイズ選択工程の前に、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含む。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含む。場合によっては、サイズ選択されたライブラリーは、約350bp～約1000bpのサイズである。場合によっては、サイズ選択ライブラリーは、約100bp～約2500bpのサイズ、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000bp～約1500bpおよび約2000bp、約2000bp～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpである。

【0203】

本明細書的全細胞または全核の消化を含む方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせを用いて行うことができる。

【0204】

別の態様では、本明細書的全細胞または全核の消化を含む方法は、QC値を得るために複数の選択されたセグメントをさらに分析する工程を含み得る。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指数(GDI)から選択される。CDEは、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、CDEは、サイズ選択前の100～2500bpのサイズのセグメントの割合として計算される。場合によっては、試料は、CDE値が少なくとも65%であるときに、さらなる分析のために選択される。場合によっては、試料は、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%である場合に、さらなる分析のために選択される。CD

I は、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算される。例えば、C D I は、6 0 0 ~ 2 5 0 0 b p のサイズを有する断片対 1 0 0 ~ 6 0 0 b p のサイズを有する断片の比の対数として計算され得る。場合によっては、試料は、C D I 値が - 1 . 5 より大きくかつ 1 未満である場合に、さらなる分析のために選択される。場合によっては、C D I 値が約 - 2 より大きく約 1 . 5 未満、約 - 1 . 9 より大きく約 1 . 5 未満、約 - 1 . 8 より大きく約 1 . 5 未満、約 - 1 . 7 より大きく約 1 . 5 未満、約 - 1 . 6 より大きく約 1 . 5 未満、約 - 1 . 5 より大きく約 1 . 5 未満、約 - 1 . 4 より大きく約 1 . 5 未満、約 - 1 . 3 より大きく約 1 . 5 未満、約 - 1 . 2 より大きく約 1 . 5 未満、約 - 1 . 1 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 5 未満、約 - 1 より大きく約 1 . 5 未満、約 - 0 . 9 より大きく約 1 . 5 未満、約 - 0 . 8 より大きく約 1 . 5 未満、約 - 0 . 7 より大きく約 1 . 5 未満、約 - 0 . 6 より大きく約 1 . 5 未満、約 - 0 . 5 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 4 未満、約 - 2 より大きく約 1 . 3 未満、約 - 2 より大きく約 1 . 2 未満、約 - 2 より大きく約 1 . 1 未満、約 - 2 より大きく約 1 未満、約 - 2 より大きく約 0 . 9 未満、約 - 2 より大きく約 0 . 8 未満、約 - 2 より大きく約 0 . 7 未満、約 - 2 より大きく約 0 . 6 未満、または約 - 2 より大きく約 0 . 5 未満である場合に、試料はさらなる分析のために選択される。

10

【 0 2 0 5 】

別の態様において、本明細書的全細胞または全核の消化を含む方法は、少数の細胞または少量の核酸を含有する少量の試料において行われる。場合によっては、安定化された生体試料は 3 , 0 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体試料は 2 , 0 0 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 1 , 0 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体試料は 5 0 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 4 0 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 3 0 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 2 0 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 1 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体試料は 5 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 4 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 3 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 2 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は 1 0 , 0 0 0 個未満の細胞を含む。場合によっては、安定化された生体試料は約 1 0 , 0 0 0 個の細胞を含む。場合によっては、安定化された生体試料は 1 0 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 9 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 8 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 7 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 6 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 5 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 4 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 3 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 2 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 1 μ g 未満の D N A を含む。場合によっては、安定化された生体試料は 0 . 5 μ g 未満の D N A を含む。

20

30

40

【 0 2 0 6 】

他の態様では、本明細書的全細胞または全核の消化を含む方法は、個々のまたは単一の細胞上で実施され得る。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【 0 2 0 7 】

50

さらなる態様では、本明細書の全細胞または全核の消化を含む方法で使用される安定化された生体試料は、DNAの断片を作製するためにDNaseなどのヌクレアーゼで処理される。場合によっては、DNaseは非配列特異的である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性である。場合によっては、DNaseは二本鎖DNAに特異的である。場合によっては、DNaseは二本鎖DNAを優先的に切断する。場合によっては、DNaseは一本鎖DNAに特異的である。場合によっては、DNaseは一本鎖DNAを優先的に切断する。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseは小球菌ヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼのうちの1つ以上から選択される。場合によっては、DNaseは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

10

20

30

40

50

【0208】

さらなる態様では、本明細書の全細胞または全核の消化を含む方法で使用される安定化された生体試料は、クロスリンク剤で処理される。場合によっては、クロスリンク剤は化学的固定剤である。場合によっては、化学的固定剤は、約2.3~2.7オングストローム(Å)のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または20 Åのスペーサー長を有することができる。化学的固定剤は、約16.1 Åの長さを有するスペーサーアームを有するエチレンジアミンビス(スクシンイミジルスクシネート)(EGS)を含むことができる。化学的固定剤は、約7.7 Åの長さのスペーサーアームを有するジスクシンイミジルグルタレート(DSG)を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよびEGS、ホルムアルデヒドおよびDSG、またはホルムアルデヒド、EGS、および、DSGを含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用される。他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな(例えば、>1 kb)リード対分離距離を有するリード対の割合を増加させることができる。例えば、図7は、得られたライブラリー(消化されたDNaseおよびMNaseの両方)をホルムアルデヒド単独でクロスリンクしたものと、ホルムアルデヒドとDSGまたはEGSでクロスリンクしたものとの比較を示す。DSGは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。DSGは膜透過性であり、細胞内クロスリンクを可能にする。DSGは、いくつかの用途においてジスクシンイミジルグルタレート(DSS)と比較して、クロスリンク効率を高めることができる。EGSは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。EGSは膜透過性であり、細胞内クロスリンクを可能にする。EGSクロスリンクは、例えば、pH8.5で3~6時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGSとの可逆的クロスリンク後にその活性の60%を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、クロスリンク剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス(2-クロロエチル)エチルアミン、ビス(2-クロロエチル)メチルアミン、トリス(2-クロロエチル)アミン、イソファミド(isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、プスルファン、シスプラチン、カルボプラチン、シシクロプラチン(cycloplatin)、エプタプラチン(eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン

、サトラブラチン、四硝酸トリブラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (mitozolomide)、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンである。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。

【0209】

さらなる態様では、本明細書の全細胞または全核の消化を含む方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0210】

さらなる態様では、本明細書の全細胞または全核の消化を含む方法は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を *fish-in* すること、および平滑末端をライゲーションすることを含む。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約15～約18ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約25、約30、約35、約40、約45、または約50ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含むことができる。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含むことができる。いくつかの実施形態では、架橋オリゴヌクレオチドは、互いに結合された複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合され (*coupled* または *linked*) 得る。場合によっては、結合された架橋オリゴヌクレオチドは、抗体が結合される試料核酸中の位置に送達され得る。

【0211】

分割およびプーリングアプローチを用いて、固有のバーコードを有する架橋オリゴヌクレオチドを生成することができる。試料の集団は、複数の群に分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドバーコードが群間で異なるが、群内では同じであるように試料に結合させることができ、試料の群を再度プールすることができ、このプロセスを複数回繰り返すことができる。このプロセスを反復すると、最終的には、集団中の各試料が固有の一連の架橋オリゴヌクレオチドバーコードを有することになり、単一試料 (例えば、単細胞、単核、単染色体) の分析が可能になる。1つの例示的な例では、ビーズの固体支持体に結合したクロスリンクした消化核の試料を8本のチューブに分け、各々が、ライゲーションされる二本鎖DNA (*dsDNA*) アダプターを含む第1のアダプター群 (第1の反復) の8つの固有のメンバーのうちの1つを含有する。8つのアダプターの各々は、核中のクロスリンクしたクロマチン凝集体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有することができるが、そうでなければ固有の *dsDNA* 配列を有する。第1のアダプター群をライゲーションした後、核をプールして戻し、洗浄してライゲーション反応成分を除去することができる。分配、ライゲーション、およびプールのスキームは、さらに2回繰り返すことができる (2回の反復)。各アダプター群からのメンバーのライゲーション後に、クロスリンクしたクロマチン凝集体を複数のバーコードに連続して結合させることができる。場合によっては、複数のアダプター群の複数のメンバーの連続ライゲーション (反復) は、バーコードの組み合わせをもた

10

20

30

40

50

らす。利用可能なバーコードの組み合わせの数は、反復当たりの群の数、および使用されるバーコードオリゴヌクレオチドの総数に依存する。例えば、それぞれ8つのメンバーを含む3つの反復は、 8^3 の可能な組み合わせを有することができる。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加または減少させることによって、および／または、反復の数を増加または減少させることによって調整することができる。1を超えるアダプター群が使用される場合、反復アダプター結合のために、分配、結合、およびプールのスキームを使用することができる。場合によっては、分配、結合、およびプールのスキームは、少なくとも3、4、5、6、7、8、9、または10回さらに繰り返され得る。場合によっては、最後のアダプター群のメンバーは、例えば、PCR増幅による配列決定ライブラリー調製中に、アダプターに結合したDNAのその後の濃縮のための配列を含む。

10

【0212】

さらなる態様では、本明細書の全細胞または全核の消化を含む方法は、せん断工程を含まない。

【0213】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1のリード対を生成するために、接合部の各側で少なくともいくつかの配列を得る工程を含む。例えば、方法は、第1のリード対を生成するために、接合部の各側で少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

20

【0214】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とを含む。

【0215】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含む。

30

【0216】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセット中のバリエーションをフェーズに割り当てる工程とを含む。

【0217】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(1)安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2)バリエーションの存在に基づいて薬物を選択すること、または、(3)安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を実施する工程とを含む。

40

【0218】

低核酸入力要件を有するHi-C方法

少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料をDNaseと接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程であって、安定化された生体試料は3,000,000個未満の細胞または10μg未満のDNAを含む方法が本明細書でさらに提供される。場合によっては、安定化された生体試料は3,000,000個未満の細胞を含む。場合によっては、安定化された生体試料は2,000,000個未満の細胞を含む。

50

50

50

本明細書の方法の様々な態様では、安定化された試料は核を含むことがある。場合によっては、安定化された試料は、50,000個以下の核を含む。場合によっては、試料は、40,000個以下の核を含む。場合によっては、試料は、30,000個以下の核を含む。場合によっては、試料は、20,000個以下の核を含む。場合によっては、試料は少なくとも10,000個の核を含む。場合によっては、試料は少なくとも20,000個の核を含む。場合によっては、試料は少なくとも30,000個の核を含む。場合によっては、試料は少なくとも40,000個の核を含む。場合によっては、試料は約10,000個の核～約50,000個の核を含む。場合によっては、試料は約20,000個の核～約50,000個の核を含む。場合によっては、試料は約30,000個の核～約50,000個の核を含む。場合によっては、試料は約40,000個の核～約50,000個の核を含む。場合によっては、試料は約10,000個の核～約40,000個の核を含む。場合によっては、試料は約10,000個の核～約30,000個の核を含む。場合によっては、試料は約10,000個の核～約20,000個の核を含む。場合によっては、試料は約20,000個の核～約50,000個の核を含む。場合によっては、試料は約20,000個の核～約40,000個の核を含む。場合によっては、試料

は約20,000個の核～約30,000個の核を含む。場合によっては、試料は約30,000個の核～約50,000個の核を含む。場合によっては、試料は約30,000個の核～約40,000個の核を含む。

【0220】

他の態様では、本明細書の低核酸入力要件を有する方法は、個々のまたは単一の細胞上で実施され得る。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

10

【0221】

他の態様では、本明細書の低核酸入力要件を有する方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程を含み得る。場合によっては、複数の選択されたセグメントは約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100bp～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpである。

20

【0222】

本明細書に提供される低核酸入力要件を有する方法の別の態様では、方法は、サイズ選択工程の前に、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含む。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含む。場合によっては、サイズ選択されたライブラリーは、約350bp～約1000bpのサイズである。場合によっては、サイズ選択ライブラリーは、約100bp～約2500bpのサイズ、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000bp～約1500bpおよび約2000bp、約2000bp～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpである。

30

40

【0223】

本明細書の低核酸入力要件を有する方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせを用いて行うことができる。

【0224】

別の態様では、本明細書の低核酸入力要件を有する方法は、QC値を得るために複数の選択されたセグメントを分析する工程をさらに含み得る。場合によっては、QC値は、クロマチン消化効率（CDE）およびクロマチン消化指数（GDI）から選択される。CDEは、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては

50

、C D E は、サイズ選択前の 1 0 0 ~ 2 5 0 0 b p のサイズのセグメントの割合として計算される。場合によっては、試料は、C D E 値が少なくとも 6 5 % であるときに、さらなる分析のために選択される。場合によっては、試料は、C D E 値が少なくとも約 5 0 %、少なくとも約 5 5 %、少なくとも約 6 0 %、少なくとも約 6 5 %、少なくとも約 7 0 %、少なくとも約 7 5 %、少なくとも約 8 0 %、少なくとも約 8 5 %、少なくとも約 9 0 %、または少なくとも約 9 5 % である場合に、さらなる分析のために選択される。C D I は、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算される。例えば、C D I は、6 0 0 ~ 2 5 0 0 b p のサイズを有する断片対 1 0 0 ~ 6 0 0 b p のサイズを有する断片の比の対数として計算され得る。場合によっては、試料は、C D I 値が - 1 . 5 より大きくかつ 1 未満である場合に、さらなる分析のために選択される。場合によっては、C D I 値が約 - 2 より大きく約 1 . 5 未満、約 - 1 . 9 より大きく約 1 . 5 未満、約 - 1 . 8 より大きく約 1 . 5 未満、約 - 1 . 7 より大きく約 1 . 5 未満、約 - 1 . 6 より大きく約 1 . 5 未満、約 - 1 . 5 より大きく約 1 . 5 未満、約 - 1 . 4 より大きく約 1 . 5 未満、約 - 1 . 3 より大きく約 1 . 5 未満、約 - 1 . 2 より大きく約 1 . 5 未満、約 - 1 . 1 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 5 未満、約 - 1 より大きく約 1 . 5 未満、約 - 0 . 9 より大きく約 1 . 5 未満、約 - 0 . 8 より大きく約 1 . 5 未満、約 - 0 . 7 より大きく約 1 . 5 未満、約 - 0 . 6 より大きく約 1 . 5 未満、約 - 0 . 5 より大きく約 1 . 5 未満、約 - 2 より大きく約 1 . 4 未満、約 - 2 より大きく約 1 . 3 未満、約 - 2 より大きく約 1 . 2 未満、約 - 2 より大きく約 1 . 1 未満、約 - 2 より大きく約 1 未満、約 - 2 より大きく約 0 . 9 未満、約 - 2 より大きく約 0 . 8 未満、約 - 2 より大きく約 0 . 7 未満、約 - 2 より大きく約 0 . 6 未満、または約 - 2 より大きく約 0 . 5 未満である場合に、試料はさらなる分析のために選択される。

10

20

【 0 2 2 5 】

別の態様において、本明細書における低核酸入力要件を有する方法において使用される安定化された生体試料は、安定化剤で処理された生体材料を含む。場合によっては、安定化された生体試料は安定化された細胞溶解物を含む。代替的に、安定化された生体試料は安定化された無傷細胞を含む。代替的に、安定化された生体試料は安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核試料を D N a s e に接触させる工程は、無傷細胞または無傷の核の溶解の前に行われる。場合によっては、細胞および／または核は、接合部において複数のセグメントの第 1 のセグメントおよび第 2 のセグメントを結合させる前に溶解される。

30

【 0 2 2 6 】

さらなる態様では、本明細書の低核酸入力要件を有する方法で使用される安定化された生体試料は、D N A の断片を作製するために D N a s e などのヌクレアーゼで処理される。場合によっては、D N a s e は非配列特異的である。場合によっては、D N a s e は、一本鎖 D N A および二本鎖 D N A の両方に対して活性である。場合によっては、D N a s e は二本鎖 D N A に特異的である。場合によっては、D N a s e は二本鎖 D N A を優先的に切断する。場合によっては、D N a s e は一本鎖 D N A に特異的である。場合によっては、D N a s e は一本鎖 D N A を優先的に切断する。場合によっては、D N a s e は D N a s e I である。場合によっては、D N a s e は D N a s e I I である。場合によっては、D N a s e は、D N a s e I および D N a s e I I の 1 つ以上から選択される。場合によっては、D N a s e は小球菌ヌクレアーゼである。場合によっては、D N a s e は、D N a s e I、D N a s e I I、および小球菌ヌクレアーゼのうちの 1 つ以上から選択される。場合によっては、D N a s e は、プロテイン A、プロテイン G、プロテイン A / G、またはプロテイン L などの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

40

【 0 2 2 7 】

さらなる態様では、本明細書の低核酸入力要件を有する方法で使用される安定化された生体試料は、クロスリンク剤で処理される。場合によっては、クロスリンク剤は

50

化学的固定剤である。場合によっては、化学的固定剤は、約 2.3 ~ 2.7 オングストローム (Å) のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約 3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または 20 Å のスペーサー長を有することができる。化学的固定剤は、約 16.1 Å の長さを有するスペーサーアームを有するエチレングリコールビス (スクシンイミジルスクシネート) (EGS) を含むことができる。化学的固定剤は、約 7.7 Å の長さのスペーサーアームを有するジスクシンイミジルグルタレート (DSG) を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよび EGS、ホルムアルデヒドおよび DSG、またはホルムアルデヒド、EGS、および、DSG を含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用される。他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな (例えば、> 1 kb) リード対分離距離を有するリード対の割合を増加させることができる。例えば、図 7 は、得られたライブラリー (消化された DNase および MNase の両方) をホルムアルデヒド単独でクロスリンクしたものと、ホルムアルデヒドと DSG または EGS でクロスリンクしたものとの比較を示す。DSG は、両方の末端に NHS エステル反応基を有し、アミノ基 (例えば、一級アミン) に対して反応性であり得る。DSG は膜透過性であり、細胞内クロスリンクを可能にする。DSG は、いくつかの用途においてジスクシンイミジルスベレート (DSS) と比較して、クロスリンク効率を高めることができる。EGS は、両方の末端に NHS エステル反応基を有し、アミノ基 (例えば、一級アミン) に対して反応性であり得る。EGS は膜透過性であり、細胞内クロスリンクを可能にする。EGS クロスリンクは、例えば、pH 8.5 で 3 ~ 6 時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGS との可逆的クロスリンク後にその活性の 60 % を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、クロスリンク剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ベンダムスチン、ビス (2-クロロエチル) エチルアミン、ビス (2-クロロエチル) メチルアミン、トリス (2-クロロエチル) アミン、イソファミド (isofamide)、カルムスチン、ロムスチン、ストレプトゾシン、プスルファン、シスプラチン、カルボプラチン、シシクロプラチン (ciccy cloplatin)、エプタプラチン (eptaplatin)、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド (mitozolomide)、テモゾロミド、マイトマイシン C、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキシソルピシン、ダウノルピシン、エピルピシン、またはイダルピシンである。場合によっては、クロスリンク剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料である。

10

20

30

40

50

【0228】

さらなる態様では、本明細書で提供される方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0229】

さらなる態様では、本明細書で提供される低核酸入力要件を有する方法は、接合部において複数のセグメントの第 1 のセグメントおよび第 2 のセグメントを結合させる工程を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を fill-in すること、および平滑末端をライゲーションすることを含む。場合によっては、結合させる工程は、少なくとも第 1 のセグメントおよび第 2 のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、結合させる工程は、少なくとも第 1 のセグメントおよび第 2 のセグメントをバーコードに接触させることを含

む。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約 5 ヌクレオチド長～約 50 ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約 15～約 18 ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約 5、約 6、約 7、約 8、約 9、約 10、約 11、約 12、約 13、約 14、約 15、約 16、約 17、約 18、約 19、約 20、約 25、約 30、約 35、約 40、約 45、または約 50 ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含み得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含むことができる。いくつかの実施形態では、架橋オリゴヌクレオチドは、互いに結合された複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテイン A、プロテイン G、プロテイン A/G、またはプロテイン L などの免疫グロブリン結合タンパク質またはその断片に結合され (coupled または linked) 得る。場合によっては、結合された架橋オリゴヌクレオチドは、抗体が結合される試料核酸中の位置に送達され得る。

10

20

30

40

50

【0230】

分割およびプーリングアプローチを用いて、固有のバーコードを有する架橋オリゴヌクレオチドを生成することができる。試料の集団は、複数の群に分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドバーコードが群間で異なるが、群内では同じであるように試料に結合させることができ、試料の群を再度プールすることができ、このプロセスを複数回繰り返すことができる。このプロセスを反復すると、最終的には、集団中の各試料が固有の一連の架橋オリゴヌクレオチドバーコードを有することになり、単一試料 (例えば、単細胞、単核、単染色体) の分析が可能になる。1つの例示的な例では、ビーズの固体支持体に結合したクロスリンクした消化核の試料を 8 本のチューブに分け、各々が、ライゲーションされる二本鎖 DNA (dsDNA) アダプターを含む第 1 のアダプター群 (第 1 の反復) の 8 つの固有のメンバーのうちの 1 つを含有する。8 つのアダプターの各々は、核中のクロスリンクしたクロマチン凝集体の核酸末端へのライゲーションのための同じ 5' オーバーハング配列を有することができるが、そうでなければ固有の dsDNA 配列を有する。第 1 のアダプター群をライゲーションした後、核をプールして戻し、洗浄してライゲーション反応成分を除去することができる。分配、ライゲーション、およびプールのスキームは、さらに 2 回繰り返すことができる (2 回の反復)。各アダプター群からのメンバーのライゲーション後に、クロスリンクしたクロマチン凝集体を複数のバーコードに連続して結合させることができる。場合によっては、複数のアダプター群の複数のメンバーの連続ライゲーション (反復) は、バーコードの組み合わせをもたらす。利用可能なバーコードの組み合わせの数は、反復当たりの群の数、および使用されるバーコードオリゴヌクレオチドの総数に依存する。例えば、それぞれ 8 つのメンバーを含む 3 つの反復は、83 の可能な組み合わせを有することができる。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加または減少させることによって、および／または、反復の数を増加または減少させることによって調整することができる。1 を超えるアダプター群が使用される場合、反復アダプター結合のために、分配、結合、およびプールのスキームを使用することができる。場合によっては、分配、結合、およびプールのスキームは、少なくとも 3、4、5、6、7、8、9、または 10 回さらに繰り返され得る。場合によっては、最後のアダプター群のメンバーは、例えば、PCR 増幅による配列決定ライブラリー調製中に、アダプターに結合した DNA のその後の濃縮のための配列を含む。

【0231】

さらなる態様では、本明細書の低核酸入力要件を有する方法は、せん断工程を含まない。

【0232】

本明細書の低核酸入力要件を有する方法のさらなる態様では、方法は、第 1 のリード対

を生成するために、接合部の各側で少なくともいくつかの配列を得る工程を含む。例えば、方法は、第1のリード対を生成するために、接合部の各側で少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

【0233】

本明細書の低核酸入力要件を有する方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とを含む。

【0234】

本明細書の低核酸入力要件を有する方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含む。

【0235】

本明細書の低核酸入力要件を有する方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセット中のバリエーションをフェーズに割り当てる工程とを含む。

【0236】

本明細書の低核酸入力要件を有する方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(1) 安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2) バリエーションの存在に基づいて薬物を選択すること、または、(3) 安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を実施する工程とを含む。

【0237】

小球菌ヌクレアーゼ (MNase) を使用する Hi-C 方法

少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料を小球菌ヌクレアーゼ (MNase) と接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程とを含み得る方法がさらに本明細書で提供される。本明細書の方法における MNase の使用は、例えば、MNase が DNA 結合タンパク質に結合していないすべての塩基対を切断することができるため、DNA 結合タンパク質が最大で一塩基対解像度によってクロマチンに結合している場所に関する特定の情報を提供することができる。さらに、MNase 消化の使用により、三次元クロマチン構造情報を解読するための接触マップおよびトポロジーに関連するドメインの作成が可能となり得る。場合によっては、MNase は、プロテイン A、プロテイン G、プロテイン A/G、またはプロテイン L などの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。

【0238】

例えば、MNase Hi-C 方法は、約1bp、2bp、3bp、4bp、5bp、6bp、7bp、8bp、9bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1000bp、2000bp、3000bp、4000bp、5000bp、6000bp、7000bp、8000bp、9000bp、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、または100kb以下の解像度でタンパク質結合またはゲノムの接触相互作用の位置を提供することができる。場合によっては、タンパク質結合部位、タンパク質フットプリント、接触相互作用、または他の特徴は、1000bp以内、900bp以内、800bp以内、700bp以内、600bp以内、500bp以内、400bp以内、300bp以内、200bp以内、190bp以内、180bp以内、170bp以内、160bp以内、150bp以内、140bp以内

、130bp以内、120bp以内、110bp以内、100bp以内、90bp以内、80bp以内、70bp以内、60bp以内、50bp以内、40bp以内、30bp以内、20bp以内、10bp以内、9bp以内、8bp以内、7bp以内、6bp以内、5bp以内、4bp以内、3bp以内、2bp以内、または1bp以内にマッピングすることができる。

【0239】

特定の態様では、MNase消化工程を含む方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程を含み得る。場合によっては、複数の選択されたセグメントは約145～約600bpであり得る。場合によっては、複数の選択されたセグメントは約100～約2500bpであり得る。場合によっては、複数の選択されたセグメントは約100～約600bpであり得る。場合によっては、複数の選択されたセグメントは約600～約2500bpであり得る。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100bp～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpであり得る。

【0240】

本明細書に提供されるMNase消化工程を含む方法の別の態様では、方法は、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含み得る。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含み得る。場合によっては、サイズ選択されたライブラリーは、約350bp～約1000bpのサイズであり得る。場合によっては、サイズ選択ライブラリーは、約100bp～約2500bpのサイズ、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000bp～約1500bp、約2000bp～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpであり得る。

【0241】

別の態様では、本明細書で提供されるMNase消化工程を含む方法は、QC値を得るために複数のセグメントを分析する工程をさらに含むことができる。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指数(GDI)から選択され得る。CDEは、所望の長さを有するセグメントの割合として計算することができる。例えば、場合によっては、CDEは、サイズ選択前の100bp～2500bpのサイズのセグメントの割合として計算することができる。場合によっては、試料は、CDE値が少なくとも65%であるときに、さらなる分析のために選択される。場合によっては、試料は、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%である場合に、さらなる分析のために選択され得る。

【0242】

GDIは、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算することができる。例えば、GDIは、600～2500bpのサイズを有する断片対100～600bpのサイズを有する断片の

比の対数として計算され得る。場合によっては、試料は、C D I 値が -1.5 より大きくかつ 1 未満である場合に、さらなる分析のために選択され得る。場合によっては、C D I 値が約 -2 より大きく約 1.5 未満、約 -1.9 より大きく約 1.5 未満、約 -1.8 より大きく約 1.5 未満、約 -1.7 より大きく約 1.5 未満、約 -1.6 より大きく約 1.5 未満、約 -1.5 より大きく約 1.5 未満、約 -1.4 より大きく約 1.5 未満、約 -1.3 より大きく約 1.5 未満、約 -1.2 より大きく約 1.5 未満、約 -1.1 より大きく約 1.5 未満、約 -2 より大きく約 1.5 未満、約 -1 より大きく約 1.5 未満、約 -0.9 より大きく約 1.5 未満、約 -0.8 より大きく約 1.5 未満、約 -0.7 より大きく約 1.5 未満、約 -0.6 より大きく約 1.5 未満、約 -0.5 より大きく約 1.5 未満、約 -2 より大きく約 1.4 未満、約 -2 より大きく約 1.3 未満、約 -2 より大きく約 1.2 未満、約 -2 より大きく約 1.1 未満、約 -2 より大きく約 1 未満、約 -2 より大きく約 0.9 未満、約 -2 より大きく約 0.8 未満、約 -2 より大きく約 0.7 未満、約 -2 より大きく約 0.6 未満、または約 -2 より大きく約 0.5 未満である場合に、試料はさらなる分析のために選択され得る。

10

【0243】

別の態様において、本明細書で提供される M N a s e 消化工程を含む方法において使用される安定化された生体試料は、安定化剤で処理された生体材料を含む。場合によっては、安定化された生体試料は安定化された細胞溶解物を含み得る。代替的に、安定化された生体試料は安定化された無傷細胞を含み得る。代替的に、安定化された生体試料は安定化された無傷の核を含み得る。場合によっては、安定化された無傷細胞または無傷の核試料を M N a s e に接触させる工程は、無傷細胞または無傷の核の溶解の前に行われ得る。場合によっては、細胞および／または核は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる前に溶解され得る。

20

【0244】

別の態様において、本明細書の M N a s e 消化工程を含む方法は、少数の細胞または少量の核酸を含有する少量の試料において行われ得る。例えば、場合によっては、安定化された生体試料は $3,000,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $2,000,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $1,000,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $500,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $400,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $300,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $200,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $100,000$ 個未満の細胞を含み得る。場合によっては、安定化された生体試料は $50,000$ 個未満の細胞を含む。場合によっては、安定化された生体試料は $40,000$ 個未満の細胞を含む。場合によっては、安定化された生体試料は $30,000$ 個未満の細胞を含む。場合によっては、安定化された生体試料は $20,000$ 個未満の細胞を含む。場合によっては、安定化された生体試料は $10,000$ 個未満の細胞を含む。場合によっては、安定化された生体試料は約 $10,000$ 個の細胞を含む。場合によっては、安定化された生体試料は $10\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $9\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $8\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $7\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $6\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $5\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $4\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $3\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $2\mu\text{g}$ 未満の D N A を含み得る。場合によっては、安定化された生体試料は $1\mu\text{g}$ 未満の D N A を含む。場合によっては、安定化された生体試料は $0.5\mu\text{g}$ 未満の D N A を含む。

30

40

【0245】

別の態様では、本明細書の M N a s e 消化工程を含む方法は、個々の細胞または単一の

50

細胞に対して行ってもよい。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【0246】

さらなる態様では、本明細書のMNase消化工程を含む方法で使用される安定化された生体試料は、DNAの断片を作製するためにDNaseなどの追加のヌクレアーゼでさらに処理され得る。場合によっては、DNaseは非配列特異的であり得る。場合によっては、DNaseは一本鎖DNAおよび二本鎖DNAの両方に対して活性であり得る。場合によっては、DNaseは二本鎖DNAに特異的であり得る。場合によっては、DNaseは二本鎖DNAを優先的に切断し得る。場合によっては、DNaseは一本鎖DNAに特異的であり得る。場合によっては、DNaseは一本鎖DNAを優先的に切断し得る。場合によっては、DNaseはDNase Iであり得る。場合によっては、DNaseはDNase IIであり得る。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択され得る。場合によっては、DNaseは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

【0247】

さらなる態様において、MNase消化工程を含む方法において使用するための本明細書に提供される安定化された生体試料は、クロスリンク剤で処理することができる。場合によっては、クロスリンク剤は化学的固定剤であり得る。場合によっては、化学的固定剤は、約2.3~2.7オングストローム(Å)のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約3Å、4Å、5Å、6Å、7Å、8Å、9Å、10Å、11Å、12Å、13Å、14Å、15Å、16Å、17Å、18Å、19Å、または20Åのスペーサー長を有することができる。化学的固定剤は、約16.1Åの長さを有するスペーサーアームを有するエチレンジアミンビス(スクシンイミジルスクシネート)(EGS)を含むことができる。化学的固定剤は、約7.7Åの長さのスペーサーアームを有するジスクシンイミジルグルタレート(DSG)を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよびEGS、ホルムアルデヒドおよびDSG、またはホルムアルデヒド、EGS、および、DSGを含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用される。他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな(例えば、>1kb)リード対分離距離を有するリード対の割合を増加させることができる。例えば、図7は、得られたライブラリー(消化されたDNaseおよびMNaseの両方)をホルムアルデヒド単独でクロスリンクしたものと、ホルムアルデヒドとDSGまたはEGSでクロスリンクしたものとの比較を示す。DSGは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。DSGは膜透過性であり、細胞内クロスリンクを可能にする。DSGは、いくつかの用途においてジスクシンイミジルスベレート(DSS)と比較して、クロスリンク効率を高めることができる。EGSは、両方の末端にNHSEステル反応基を有し、アミノ基(例えば、一級アミン)に対して反応性であり得る。EGSは膜透過性であり、細胞内クロスリンクを可能にする。EGSクロスリンクは、例えば、pH8.5で3~6時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGSとの可逆的クロスリンク後にその活性の60%を保持した。場合によっては、化学的固定剤はソラレンを含み得る。場合によっては、

クロスリンキング剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン、メルファラン、ペンダムスチン、ビス（２－クロロエチル）エチルアミン、ビス（２－クロロエチル）メチルアミン、トリス（２－クロロエチル）アミン、イソファミド（i s o f a m i d e）、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン（c i c y c l o p l a t i n）、エプタプラチン（e p t a p l a t i n）、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド（m i t o z o l o m i d e）、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキシソルピシン、ダウノルピシン、エピルピシン、またはイダルピシンであり得る。場合によっては、クロスリンキング剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料であり得る。

10

【0248】

さらなる態様では、本明細書で提供されるM N a s e消化工程を含む方法は、複数の選択されたセグメントを抗体に接触させる工程を含み得る。場合によっては、オリゴヌクレオチドアダプターにテザリングされた免疫グロブリン結合タンパク質またはその断片は、複数の選択されたセグメントに結合した抗体にテザリングされ得る。

【0249】

さらなる態様では、本明細書で提供されるM N a s e消化工程を含む方法は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程を含む。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端をf i l l i n gすること、および平滑末端をライゲーションすることを含み得る。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを架橋オリゴヌクレオチドに接触させることを含み得る。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含み得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約15～約18ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約25、約30、約35、約40、約45、または約50ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含み得る。

20

30

【0250】

本明細書のM N a s e消化工程を含む方法のさらなる態様では、方法は、第1のリード対を生成するために、接合部の各側で少なくともいくつかの配列を得る工程を含むことができる。例えば、方法は、第1のリード対を生成するために、接合部の各側で少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

40

【0251】

本明細書におけるM N a s e消化工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングすることと、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定することとを含むことができる。

【0252】

本明細書のM N a s e消化工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含むことができる。

50

【0253】

本明細書におけるMNase消化工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセット中のバリエーションをフェーズに割り当てる工程とを含む。

【0254】

本明細書のMNase消化工程を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(1)安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2)バリエーションの存在に基づいて薬物を選択すること、または、(3)安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を実施する工程とを含むことができる。

10

【0255】

HiChIP、HiChIRP、およびメチルHiCのための改善された方法

HiChIPは、HiCの方法をクロマチン免疫沈降の方法と組み合わせるアプローチであり、関心対象の1つ以上のタンパク質を含む相互作用の標的分析を可能にする。近接ライゲーションされた核酸を調製することができ、標的領域をさらなる分析のために免疫沈降させることができる。関連するアプローチであるHiChIRPは、RNA精製(CHiRP)濃縮によるクロマチン単離をHiC法と組み合わせて使用し、長い非コードRNA(lncRNA)の足場機能などのRNAの調査を可能にする。メチル-HiCは、メチル化分析をHiC法と組み合わせ、染色体立体配座およびDNAメチローム情報の同時捕捉を可能にする。メチル-HiCは、核内で空間的に近接している遠位ゲノムセグメント間の同調的なDNAメチル化状態を明らかにし、混合集団におけるクロマチン構造およびDNAメチロームの両方の異質性を描写し、複雑な組織における細胞型特異的クロマチン構成およびエピゲノムの同時の特徴付けを可能にすることができる。これらの方法および他の方法は、限定されないが、サイズ選択工程、表面結合工程(例えば、SPRIBIーズなどのビーズへの結合)、近接ライゲーションを行うための架橋オリゴヌクレオチドの使用、近接ライゲーションを行うための組換えの使用などを含む本開示の技法の使用によって改善することができる。

20

【0256】

さらなる態様では、例えば、核酸結合タンパク質に結合した核酸の免疫沈降によって、またはメチル化核酸の免疫沈降によって、少なくとも1つの核酸結合タンパク質と複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料をDNaseと接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程と、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程と、を含むことができる。HiChIP、HiChIRP、およびメチルHiCのための改善された方法が本明細書で提供される。代替的に、または、組み合わせで、本明細書の方法は、例えば、核酸結合タンパク質に結合した核酸の免疫沈降によって、またはメチル化核酸の免疫沈降によって、少なくとも1つの核酸結合タンパク質と複合体化された核酸分子を含む安定化された生体試料を得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体試料を小球菌ヌクレアーゼ(MNase)と接触させる工程と、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程とを含むことができる。

30

40

【0257】

本明細書のHiChIP、HiChIRP、およびメチルHiCのための改善された方法のいくつかの態様では、安定化された生体試料は無傷細胞および／または無傷の核を含むことができる。場合によっては、安定化された生体試料は安定化された無傷細胞を含むことができる。代替的に、または、組み合わせで、安定化された生体試料は安定化された無傷の核を含むことができる。場合によっては、安定化された無傷細胞または無傷の核試料をDNaseに接触させる工程は、無傷細胞または無傷の核の溶解の前に行われ得る。

50

場合によっては、細胞および／または核は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる前に溶解され得る。

【0258】

他の態様では、本明細書のHiChIP、HiChIRP、およびメチルHiCの改善された方法を含む方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択に供する工程を含むことができる。場合によっては、複数の選択されたセグメントは約145～約600bpであり得る。場合によっては、複数の選択されたセグメントは約100～約2500bpであり得る。場合によっては、複数の選択されたセグメントは約100～約600bpであり得る。場合によっては、複数の選択されたセグメントは約600～約2500bpであり得る。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100bp～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpであり得る。

10

【0259】

本明細書のHiChIP、HiChIRP、およびメチルHiCの改善された方法を含む方法の別の態様では、方法は、サイズ選択工程の前に、複数のセグメントから配列決定ライブラリーを調製する工程をさらに含み得る。いくつかの実施形態において、方法は、サイズ選択ライブラリーを得るために、配列決定ライブラリーをサイズ選択に供する工程をさらに含み得る。場合によっては、サイズ選択されたライブラリーは、約350bp～約1000bpのサイズであり得る。場合によっては、サイズ選択ライブラリーは、約100bp～約2500bpのサイズ、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000bp～約1500bp、約2000bp～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpであり得る。

20

30

【0260】

本明細書のHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、それらの組み合わせ、または任意の他の適切な方法を用いて行うことができる。

【0261】

別の態様において、本明細書のHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法は、QC値を得るために複数の選択されたセグメントをさらに分析する工程を含み得る。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指数(GDI)から選択され得る。CDEは、所望の長さを有するセグメントの割合として計算することができる。例えば、場合によっては、CDEは、サイズ選択前の100～2500bpのサイズのセグメントの割合として計算することができる。場合によっては、試料は、CDE値が少なくとも65%であるときに、さらなる分析のために選択される。場合によっては、試料は、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%である場合に、さらなる分析のために選択され得る。

40

50

【0262】

C D I は、サイズ選択前の、モノヌクレオソームサイズのセグメントの数対ジヌクレオソームサイズのセグメントの数の比として計算することができる。例えば、C D I は、600～2500bpのサイズを有する断片対100～600bpのサイズを有する断片の比の対数として計算され得る。場合によっては、試料は、C D I 値が-1.5より大きくかつ1未満である場合に、さらなる分析のために選択され得る。場合によっては、C D I 値が約-2より大きく約1.5未満、約-1.9より大きく約1.5未満、約-1.8より大きく約1.5未満、約-1.7より大きく約1.5未満、約-1.6より大きく約1.5未満、約-1.5より大きく約1.5未満、約-1.4より大きく約1.5未満、約-1.3より大きく約1.5未満、約-1.2より大きく約1.5未満、約-1.1より大きく約1.5未満、約-2より大きく約1.5未満、約-1より大きく約1.5未満、約-0.9より大きく約1.5未満、約-0.8より大きく約1.5未満、約-0.7より大きく約1.5未満、約-0.6より大きく約1.5未満、約-0.5より大きく約1.5未満、約-2より大きく約1.4未満、約-2より大きく約1.3未満、約-2より大きく約1.2未満、約-2より大きく約1.1未満、約-2より大きく約1未満、約-2より大きく約0.9未満、約-2より大きく約0.8未満、約-2より大きく約0.7未満、約-2より大きく約0.6未満、または約-2より大きく約0.5未満である場合に、試料はさらなる分析のために選択され得る。

10

【0263】

他の態様では、本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、少数の細胞または少量の核酸を含有する少量の試料において行うことができる。場合によっては、安定化された生体試料は3,000,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は2,000,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は1,000,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は500,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は400,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は300,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は200,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は100,000個未満の細胞を含み得る。場合によっては、安定化された生体試料は50,000個未満の細胞を含む。場合によっては、安定化された生体試料は40,000個未満の細胞を含む。場合によっては、安定化された生体試料は30,000個未満の細胞を含む。場合によっては、安定化された生体試料は20,000個未満の細胞を含む。場合によっては、安定化された生体試料は10,000個の細胞を含む。場合によっては、安定化された生体試料は10μg未満のDNAを含み得る。場合によっては、安定化された生体試料は9μg未満のDNAを含み得る。場合によっては、安定化された生体試料は8μg未満のDNAを含み得る。場合によっては、安定化された生体試料は7μg未満のDNAを含み得る。場合によっては、安定化された生体試料は6μg未満のDNAを含み得る。場合によっては、安定化された生体試料は5μg未満のDNAを含み得る。場合によっては、安定化された生体試料は4μg未満のDNAを含み得る。場合によっては、安定化された生体試料は3μg未満のDNAを含み得る。場合によっては、安定化された生体試料は2μg未満のDNAを含み得る。場合によっては、安定化された生体試料は1μg未満のDNAを含み得る。場合によっては、安定化された生体試料は0.5μg未満のDNAを含み得る。

20

30

40

【0264】

他の態様では、本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、個々の細胞または単一の細胞に対して行ってもよい。例えば、本明細書の方法は、個々のパーティションに分配された細胞に対して行うことができる。例示的なパーティションとしては、限定されないが、ウェル、エマルジョン中の液滴、または本明細書の他の箇所に記載されるような差次的に配列決定されたリンカー分子

50

の別個のパッチを含む表面位置（例えば、アレイスポット、ビーズなど）が挙げられる。追加のパーティションも企図され、本明細書に開示される方法、組成物、およびシステムと一貫している。

【0265】

さらなる態様では、本明細書のHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法は、DNAの断片を作製するためにDNaseなどのヌクレアーゼで処理することができる。場合によっては、DNaseは非配列特異的であり得る。場合によっては、DNaseは一本鎖DNAおよび二本鎖DNAの両方に対して活性であり得る。場合によっては、DNaseは二本鎖DNAに特異的であり得る。場合によっては、DNaseは二本鎖DNAを優先的に切断し得る。場合によっては、DNaseは一本鎖DNAに特異的であり得る。場合によっては、DNaseは一本鎖DNAを優先的に切断し得る。場合によっては、DNaseはDNase Iであり得る。場合によっては、DNaseはDNase IIであり得る。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択され得る。場合によっては、DNaseは小球菌ヌクレアーゼであり得る。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼのうちの1つ以上から選択され得る。場合によっては、DNaseは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合または融合され得る。他の適切なヌクレアーゼも本開示の範囲内である。

【0266】

さらなる態様では、本明細書のHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法は、クロスリンク剤で処理され得る。場合によっては、クロスリンク剤は化学的固定剤であり得る。場合によっては、化学的固定剤は、約2.3〜2.7オングストローム（Å）のスペーサーアーム長を有するホルムアルデヒドを含んでいる。場合によっては、化学的固定剤は、長いスペーサーアーム長を有するクロスリンク剤を含み、例えば、クロスリンク剤は、少なくとも約3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または20 Åのスペーサー長を有することができる。化学的固定剤は、約16.1 Åの長さを有するスペーサーアームを有するエチレングリコールビス（スクシンイミジルスクシネート）（EGS）を含むことができる。化学的固定剤は、約7.7 Åの長さのスペーサーアームを有するジスクシンイミジルグルタレート（DSG）を含むことができる。場合によっては、化学的固定剤は、ホルムアルデヒドおよびEGS、ホルムアルデヒドおよびDSG、またはホルムアルデヒド、EGS、および、DSGを含む。複数の化学的固定剤が使用される場合によっては、各化学的固定剤は連続して使用される。他の場合には、複数の化学的固定剤のいくつかまたはすべてが、同時に試料に適用される。長いスペーサーアームを有するクロスリンカーの使用は、大きな（例えば、>1 kb）リード対分離距離を有するリード対の割合を増加させることができる。例えば、図7は、得られたライブラリー（消化されたDNaseおよびMNaseの両方）をホルムアルデヒド単独でクロスリンクしたものと、ホルムアルデヒドとDSGまたはEGSでクロスリンクしたものとを比較を示す。DSGは、両方の末端にNHSEステル反応基を有し、アミノ基（例えば、一級アミン）に対して反応性であり得る。DSGは膜透過性であり、細胞内クロスリンクを可能にする。DSGは、いくつかの用途においてジスクシンイミジルスベレート（DSS）と比較して、クロスリンク効率を高めることができる。EGSは、両方の末端にNHSEステル反応基を有し、アミノ基（例えば、一級アミン）に対して反応性であり得る。EGSは膜透過性であり、細胞内クロスリンクを可能にする。EGSクロスリンクは、例えば、pH 8.5で3〜6時間ヒドロキシルアミンで処理することによって脱クロスリンクすることができる。例えば、ラクトースデヒドロゲナーゼは、EGSとの可逆的クロスリンク後にその活性の60%を保持した。場合によっては、化学的固定剤はソラレンを含み得る。場合によっては、クロスリンク剤は、紫外線、クロルメチン、シクロホスファミド、クロラムブシル、ウラムスチン

10

20

30

40

50

、メルファラン、ベンダムスチン、ビス（２－クロロエチル）エチルアミン、ビス（２－クロロエチル）メチルアミン、トリス（２－クロロエチル）アミン、イソファミド（i s o f a m i d e）、カルムスチン、ロムスチン、ストレプトゾシン、ブスルファン、シスプラチン、カルボプラチン、シシクロプラチン（c i c y c l o p l a t i n）、エプタプラチン（e p t a p l a t i n）、ロバプラチン、ミリプラチン、ネダプラチン、オキサリプラチン、ピコプラチン、サトラプラチン、四硝酸トリプラチン、プロカルバジン、アルトレタミン、ダカルバジン、ミトゾロミド（m i t o z o l o m i d e）、テモゾロミド、マイトマイシンC、亜硝酸、ホルムアルデヒド、アセチルアルデヒド、ドキソルビシン、ダウノルビシン、エピルビシン、またはイダルビシンであり得る。場合によっては、クロスリンキング剤は挿入剤、抗生物質、または副溝結合剤を含む。場合によっては、安定化された生体試料はクロスリンクしたパラフィン包埋した組織試料であり得る。

10

【0267】

さらなる態様では、本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程を含み得る。場合によっては、結合させる工程は、バイオチンタグ付きヌクレオチドを使用して粘着末端をf i l l e r i nすること、および平滑末端をライゲーションすることを含むことができる。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを架橋オリゴヌクレオチドに接触させることを含むことができる。場合によっては、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含むことができる。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。いくつかの実施形態では、本明細書の架橋オリゴヌクレオチドは、約15～約18ヌクレオチド長であり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約25、約30、約35、約40、約45、または約50ヌクレオチド長であり得る。いくつかの実施形態において、本明細書の架橋オリゴヌクレオチドは、バーコードを含み得る。

20

【0268】

さらなる態様では、本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、せん断工程を含まない。

30

【0269】

本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリード対を生成するために、接合部の各側で少なくともいくつかの配列を得る工程を含み得る。例えば、方法は、第1のリード対を生成するために、接合部の各側で少なくとも約50 b p、少なくとも約100 b p、少なくとも約150 b p、少なくとも約200 b p、少なくとも約250 b p、または少なくとも約300 b pの配列を得る工程を含み得る。

【0270】

本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、ゲノムに対する順序および／または配向を表すコンティグのセットを通る経路を決定する工程とを含み得る。

40

【0271】

本明細書のH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体試料中の構造バリエーションの存在またはヘテロ接合性の喪失を決定する工程とを含み得る。

【0272】

本明細書の全細胞または全核の消化を含む方法のさらなる態様では、方法は、第1の

50

ード対をコンティグのセットにマッピングする工程と、コンティグのセット中のバリエーションをフェーズに割り当てる工程とを含み得る。

【0273】

本明細書の H i C h I P、H i C h I R P、およびメチル H i C のための改善された方法を含む方法のさらなる態様では、方法は、第1のリード対をコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセットにおけるバリエーションの存在を決定する工程と、(1)安定化された生体試料についての疾患段階、予後、または処置の経過を同定すること、(2)バリエーションの存在に基づいて薬物を選択すること、または、(3)安定化された生体試料についての薬物効力を同定することの1つ以上から選択される工程を実施する工程とを含み得る。

10

【0274】

ロングレンジのリード対の生成

本開示は、極めてロングレンジのリード対を生成し、そのデータを前述の追跡のすべての進歩のために利用するための方法を提供する。いくつかの実施形態では、本開示は、約3億リード対のみを有する高度に連続した正確なヒトゲノムアセンブリを生成する方法を提供する。他の実施形態では、本開示は、99%以上の精度でヒトゲノム中のヘテロ接合性バリエーションの90%以上をフェージングする方法を提供する。さらに、本開示によって生成されるリード対の範囲は、はるかに大きいゲノム距離に及ぶように拡張することができる。アセンブリは、極めてロングレンジのリード対ライブラリーに加えて、標準的なショットガンライブラリーから作製される。さらに他の実施形態において、本開示は、配列決定データのこれらのセットの両方を利用することができるソフトウェアを提供する。フェージングされたバリエーションは、単一のロングレンジのリード対ライブラリーを用いて生成され、そこからのリードは、参照ゲノムにマッピングされ、その後、個体の2つの親染色体のうちの1つにバリエーションを割り当てるために使用される。最後に、本開示は、例外的に長いロングリードを生成するように、公知の技術を使用してさらに大きなDNA断片を抽出することをもたらす。

20

【0275】

これらの反復がアセンブリおよびアラインメントのプロセスを妨げるメカニズムはかなり簡単であり、最終的には曖昧さの結果である。大きな反復領域の場合、問題はスパンの1つであり得る。リードまたはリード対が反復領域に及ぶほど十分に長くない場合、反復要素に隣接する領域を信頼できる形で結合することができない場合がある。小さな反復要素の場合、問題は主に配置であり得る。ある領域がゲノムにおいて共通する2つの反復要素と隣接している場合、その正確な配置を決定することは、不可能ではないにしても、隣接する要素がそれらのクラスの他のすべてに類似しているため、困難になる。両方の場合において、同定、したがって、特定の反復配列の配置を困難になるのは、反復配列中に区別する情報が欠如しているからである。必要とされているのは、反復領域によって囲まれたまたは分離された固有のセグメント間の結合を実験的に確立する能力である。

30

【0276】

本開示の方法は、これらの反復領域によってもたらされる実質的な障壁を克服することによってゲノムの分野を進歩させ、それによってゲノム分析の多くの領域における重要な進歩を可能にすることができる。以前の技術を用いて *de novo* アセンブリを実施するためには、多くの小さな足場に断片化されたアセンブリを準備するか、または、大きな挿入ライブラリーを生成するための、もしくはより連続したアセンブリを生成するための他のアプローチを使用するために、かなりの時間および資源を必要とするかのいずれかではない。そのようなアプローチは、非常に深いシーケンシングカバレッジの獲得、BACまたはホスミドライブラリーの構築、光学マッピング、または、これらおよび/もしくは他の技術のいくつかの組み合わせを含み得る。資源および時間の厳しい要件のせいで、そのようなアプローチが大部分の小規模研究室に浸透せず、非モデル生物の研究が妨げられている。本明細書に記載される方法は、非常にロングレンジのリード対を生成することができるため、デノボアセンブリは、単一の配列決定の実行で達成することがで

40

50

きる。これにより、組立コストが桁違いに削減され、必要とされる時間が数ヶ月または数年から数週間にまで短縮されることになる。場合によっては、本明細書に開示される方法は、14日未満、13日未満、12日未満、11日未満、10日未満、9日未満、8日未満、7日未満、6日未満、5日未満、4日未満、または前述の指定された期間のいずれか2つの間の範囲内で、複数のリード対を生成することを可能にする。例えば、方法は、約10日～14日で複数のリード対を生成することを可能にすることができる。生物の大部分のニッチについてさえゲノムを構築することは日常的となり、系統発生分析は比較の欠如に悩まされることなく、Genome 10kなどのプロジェクトを実現することができる。

【0277】

同様に、医療目的のための構造分析およびフェージング分析にも課題が残っている。癌、同じ種類の癌を持つ個体、または同じ腫瘍の中でさえ、驚くべき異種性が存在する。結果として生じる効果から原因となるものを引き出すことは、試料あたりの低いコストで非常に高い精度およびスループットを必要とする。オーダーメイド医療の領域では、ゲノムケアの絶対的基準のうち1つは、大小の構造再編成および新規な突然変異を含む、徹底的に特徴付けられかつフェージングされたすべてのバリエーションを持つ、配列決定されたゲノムである。以前の技術でこれを達成するためには、現在ではあまりに高価でありかつ慣例的な医療処置を要求する、デノボアセンブリに必要なものと同様の労力を必要とする。開示された方法は、低コストで完全に正確なゲノムを迅速に産生することができ、それによって、ヒト疾患の研究および処置において多くの高度に求められる能力をもたらすことができる。

【0278】

本明細書に開示される方法をフェージングに適用することは、統計的アプローチの利便性を、家族分析の精度と組み合わせることができ、いずれかの方法を単独で使用するよりも大きな節約—費用、労力、および試料—をもたらす。以前の技術では禁止的である非常に望ましいフェージング分析である新規バリエーションフェージング解析は、本明細書に開示される方法を使用して容易に実施することができる。このことは、ヒトの変種の大部分が稀であるため（5%未満のマイナー対立遺伝子頻度）、特に重要である。フェージング情報は、非連鎖遺伝子型と比較して、高度に結合したハプロタイプのネットワーク（単一染色体に割り当てられたバリエーションの集合）から有意な利点を得る集団遺伝子研究にとって貴重である。ハプロタイプ情報は、集団のサイズ、移動、および亜集団間の交換の歴史的変化のより高解像度の研究を可能にし得、特定のバリエーションを特定の親および祖父母に遡って追跡することを可能にする。これは、ひいては、疾患に関連するバリエーションの遺伝的伝達、および単一の個体においてまとめられた場合のバリエーション間の相互作用を明らかにする。本開示の方法は最終的には、極めてロングレンジのリード対（XLRP）ライブラリーの調製、配列決定、および分析を可能にすることができる。

【0279】

本開示のいくつかの実施形態では、対象由来の組織またはDNA試料が提供され得、本方法は、組み立てられたゲノム、コールされたバリエーション（大きな構造バリエーションを含む）とのアラインメント、フェージングされたバリエーションコール、または任意の追加の解析を返すことができる。他の実施形態において、本明細書に開示される方法は、個体のためのXLRPライブラリーを直接提供することができる。

【0280】

極めてロングレンジのリード対

本開示の様々な実施形態において、本明細書に開示される方法は、長い距離分離された極めてロングレンジのリード対を生成することができる。この距離の上限は、大きなサイズのDNA試料を集める能力により改善され得る。場合によっては、リード対は、最大で50、60、70、80、90、100、125、150、175、200、225、250、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000kbp、またはそれ以上のゲノム

距離におよぶ。いくつかの例において、リード対は、最大500kbpのゲノム距離におよぶことができる。他の例において、リード対は、最大2000kbpのゲノム距離におよぶことができる。本明細書に開示される方法は、分子生物学の標準技術に基づいて組み込みかつ構築することができ、さらに効率、特異性、およびゲノムカバレッジの増大に十分に適している。場合によっては、リード対は、約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、60、または90日未満で生成することができる。いくつかの例では、リード対は約14日未満で生成することができる。さらなる例では、リード対は約10日未満で生成することができる。場合によっては、本開示の方法は、複数のコンティグの正確な順序付けおよび／または配向において、少なくとも約50%、約60%、約70%、約80%、約90%、約95%、約99%、または約100%の精度で、リード対の約5%、約10%、約15%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%、約95%、約99%、または約100%超を提供することができる。例えば、方法は、複数のコンティグの正確な順序付けおよび／または配向において約90～100%の精度を提供することができる。

10

20

30

40

50

【0281】

他の実施形態において、本明細書に開示される方法は、現在利用される配列決定技術と共に使用することができる。例えば、この方法は、十分に試験されおよび／または広く配備された配列決定機器と組み合わせて使用することができる。さらなる実施形態において、本明細書に開示される方法は、現在利用される配列決定技術に由来する技術と手法と共に使用することができる。

【0282】

本開示の方法は、多種多様な生物のためのデノボゲノムアセンブリを劇的に単純化することができる。以前の技術を使用して、そのようなアセンブリは、現在、経済的なメイトペアライブラリーの短い挿入物によって制限されている。フォスミドでアクセス可能な最大40～50kbpのゲノム距離でリード対を生成することが可能であり得るが、これらは高価であり、扱いにくく、一ヒトでは300kbp～5Mbpのサイズの範囲であり得るセントロメア内のものを含む最も長い反復ストレッチに及ぶには短すぎる。場合によっては、本明細書に開示される方法は、長い距離（例えば、メガベース以上）に及ぶことが可能なリード対を提供し、それによって、これらの足場完全性の課題を克服することができる。したがって、染色体レベルのアセンブリの産生は、本開示の方法を利用することによって日常的になり得る。アセンブリのためのより労力を要する手段—現在、研究研究室では信じられない量の時間および金銭がかかっており、広範なゲノムカタログを妨げている—は不要になり可能性があり、より有意義な分析のために資源が解放される。同様に、ロングレンジフェージング情報の獲得は、集団のゲノム研究、系統発生研究、および疾患研究にさらなる膨大な力を提供することができる。本明細書に開示される方法は、多数の個体についての正確なフェージングを可能にし、したがって、集団および深い時間レベルでゲノムをプローブする我々の能力の幅および深さを拡張する。

【0283】

個別化医療の分野では、本明細書に開示される方法から生成されるXLRPリード対は、正確で、低コストで、フェージングされた、迅速に産生される個人ゲノムへの有意義な進歩を表している。現在の方法は、長距離でバリエーションをフェージングする能力が不十分であり、それによって、化合物ヘテロ接合遺伝子型の表現型の影響の特徴付けを妨げる。さらに、ゲノム疾患に実質的な関心対象の構造バリエーションは、それらを研究するために使用されるリードおよびリード対挿入物と比較してサイズが大きいいため、現在の技術を用いて正確に同定および特徴付けすることが困難である。数十キロベース～メガベース以上にわたるリード対は、この困難さを軽減するのに役立つことができ、それによって、構造多型の高度に並列かつ個別化された分析を可能にする。

【0284】

基本的な進化研究および生物医学研究は、ハイスループット配列決定における技術的進

歩によって推進されている。全ゲノム配列決定およびアセンブリは、大きなゲノム配列決定センターの出所となるように使用されるが、市販のシーケンサーは、現在、ほとんどの研究大学がこれらの機械の1つまたは複数を有するほど安価である。現在、大量のDNA配列データを生成することは比較的安価である。しかしながら、現在の技術を用いて高品質で高度に連続したゲノム配列を産生することは、理論上および実際には依然として困難である。さらに、ヒトを含む、分析を目指すであろうほとんどの生物は二倍体であるため、各個体はゲノムの2つの半数体コピーを有する。ヘテロ接合性の部位（例えば、母親によって与えられる対立遺伝子が父親によって与えられる対立遺伝子と異なる場合）では、どの対立遺伝子のセットがどの親から来たかを知ることは困難である（ハプロタイプフェージングとして知られる）。この情報は、疾患および形質の関連研究などのいくつかの進化的研究および生物医学的研究を実施するために使用することができる。

10

【0285】

様々な実施形態では、本開示は、DNA調製のための技術と、所定のゲノム内の短い、中程度の、および長い結合のハイスループット発見のためのペアエンド配列決定とを組み合わせる、ゲノムアセンブリのための方法を提供する。本開示はさらに、ゲノムアセンブリを補助するために、ハプロタイプフェージングのために、および／またはメタゲノム研究のために、これらの結合を使用する方法を提供する。本明細書に提示される方法は、対象のゲノムのアセンブリを決定するために使用することができるが、本明細書に提示される方法は、染色体などの対象のゲノムの部分のアセンブリ、または様々な長さの対象のクロマチンのアセンブリを決定するために使用することも理解されたい。

20

【0286】

いくつかの実施形態において、本開示は、対象から得られた標的DNAの配列決定断片から複数のコンティグを生成する工程を含む、本明細書に開示される1つ以上の方法を提供する。標的DNAの長いストレッチは、1つ以上のヌクレアーゼ（例えば、DNase I、DNase II、小球菌ヌクレアーゼなど）でDNAを切断することによって断片化することができる。複数の配列決定リードを得るために、得られた断片を、ハイスループット配列決定方法を用いて配列決定することができる。本開示の方法とともに使用されるハイスループット配列決定方法の例としては、限定されないが、Roche Diagnosticsによって開発された454パイロシークエンシング方法、Illuminaによって開発された「クラスター」配列決定方法、Life Technologiesによって開発されたSOLiD and Ion semiconductor配列決定方法、ならびにComplete Genomicsによって開発されたDNAナノボール配列決定法が挙げられる。次いで、異なる配列決定リードの重複末端をアセンブルしてコンティグを形成することができる。代替的に、断片化された標的DNAをベクターにクローニングすることができる。次いで、細胞または生物にDNAベクターをトランスフェクトしてライブラリーを形成する。トランスフェクトした細胞または生物を複製した後、ベクターを単離し、配列決定して、複数の配列決定リードを生成する。その後、異なる配列決定リードの重複末端をアセンブルしてコンティグを形成することができる。

30

【0287】

特にハイスループット配列決定技術によるゲノムアセンブリは、問題となり得る。多くの場合、アセンブリは、数千個または数万個の短いコンティグからなる。これらのコンティグの順序および配向は一般に不明であり、ゲノムアセンブリの有用性を制限している。これらの足場を順序付け、配向するための技術が存在するが、それらは概して、高価で、多くの人手を要し、しばしば、非常にロングレンジの相互作用を発見することに失敗する。

40

【0288】

コンティグを生成するために使用される標的DNAを含む試料は、体液（例えば、血液、尿、血清、リンパ液、唾液、口腔スワブ、肛門および膣の分泌物、汗、および精液など）を採取する、組織を採取すること、または細胞／生物を収集することを含む、任意の数の手段によって対象から得ることができる。得られた試料は、単一のタイプの細胞／生物

50

から構成されてもよく、または複数のタイプの細胞／生物から構成されてもよい。DNAは、対象の試料から抽出および調製することができる。例えば、試料は、公知の溶解緩衝液、超音波処理技術、エレクトロポレーションなどを使用して、ポリヌクレオチドを含む細胞を溶解するように処理され得る。アルコール抽出、セシウム勾配、および／またはカラムクロマトグラフィーを使用することによって、標的DNAをさらに精製して、タンパク質などの汚染物質を除去することができる。

【0289】

本開示の他の実施形態では、非常に高分子量のDNAを抽出する方法が提供される。場合によっては、XLRPライブラリーからのデータは、入力DNAの断片サイズを増加させることによって改善され得る。いくつかの例では、細胞からのDNAのメガベースサイズの断片の抽出は、ゲノム中のメガベースによって分離されたリード対を生成することができる。場合によっては、生成されたリード対は、約10kB、約50kB、約100kB、約200kB、約500kB、約1Mb、約2Mb、約5Mb、約10Mb、または約100Mbを超えるスパンにわたる配列情報を提供することができる。いくつかの例において、リード対は、約500kBを超えるスパンにわたる配列情報を提供することができる。さらなる例において、リード対は、約2Mbを超えるスパンにわたる配列情報を提供することができる。場合によっては、非常に高分子量のDNAは、非常に穏やかな細胞溶解 (Teague, B. et al. (2010) Proc. Nat. Acad. Sci. USA 107 (24), 10848-53) およびアガロースプラグ (Schwartz, D. C., & Cantor, C. R. (1984) Cell, 37 (1), 67-75) によって抽出することができる。他の場合では、長さが最大メガベースのDNA分子を精製することができる市販の機械を使用して、非常に高分子量のDNAを抽出することができる。

【0290】

染色体の物理的レイアウトのプロービング

様々な実施形態では、本開示は、生細胞内の染色体の物理的なレイアウトをプローブする工程を含む、本明細書に開示される1つ以上の方法を提供する。配列決定を通じて染色体の物理的なレイアウトを調査する技術の例は、染色体立体配座キャプチャ (「3C」)、環状染色体立体配座キャプチャ (「4C」)、カーボンコピー染色体キャプチャ (「5C」)、およびHi-Cベースの方法などの技術の「C」ファミリー、ならびに、ChIP-Loop、ChIP-PET、およびHiChIPなどのChIPベースの方法を含む。これらの技術は、核の中の空間的関係を固めるために生細胞のクロマチンの固定化を利用する。生成物のその後のプロセッシングおよび配列決定により、研究者は、ゲノム領域間の近接の関連性のマトリクスを回復することが可能になる。さらなる分析により、これらの関連性を使用して、染色体が生きている核に物理的に配置されときの染色体の三次元幾何学的マップを生成することができる。このような技術は、生細胞における染色体の離散的な空間的組織化を説明し、染色体遺伝子座間の機能的相互作用の正確な見解を提供する。これらの機能的な研究に支障をきたした1つの問題は、非特異的相互作用の存在、染色体の近接性以上のものに起因しないデータに存在する関連性であった。本開示では、これらの非特異的な染色体内部相互作用は、アセンブリのための貴重な情報を提供するために、本明細書に提示される方法によって捕捉される。

【0291】

いくつかの実施形態において、染色体内部相互作用は染色体の結合性と関連づけられる。場合によっては、染色体内部データがゲノムアセンブリを補助することができる。場合によっては、クロマチンはインビトロで再構成される。このことは有利であり得る。なぜなら、クロマチン、具体的には、クロマチンの主要タンパク質成分であるヒストンが、配列決定を通じてクロマチンの立体配座および構造を検出するための技術の最も一般的な「C」ファミリー：3C、4C、5C、およびHi-Cの下での固定化に重要であるからである。クロマチンは、配列に関して非常に非特異的であり、通常はゲノムにわたって均一にアセンブルする。場合によっては、クロマチンを使用しない種のゲノムは、再構成されたク

ロマチン上でアセンブルすることができ、それによって、本開示のホライゾン生命のすべてのドメインに拡張する。

【0292】

クロマチン立体構造捕捉技術を要約する。手短に言えば、物理的に近接しているゲノム領域間にクロスリンクが形成される。クロマチン内のDNA分子、例えば、ゲノムDNAへのタンパク質（ヒストンなど）のクロスリンクは、本明細書の他の箇所ですらに詳細に記載されるか、またはそうでなければ公知である適切な方法に従って達成することができる。場合によっては、2つ以上のヌクレオチド配列は、1つ以上のヌクレオチド配列に結合したタンパク質を介してクロスリンクされ得る。1つのアプローチは、クロマチンを紫外線照射に曝露することである（Gilmour et al., Proc. Nat'l. Acad. Sci. USA 81:4275-4279, 1984）。ポリヌクレオチドセグメントのクロスリンクは、化学的または物理的（例えば、光学的）なクロスリンクなどの他のアプローチを利用して実施され得る。適切な化学的クロスリンク剤としては、限定されないが、ホルムアルデヒドおよびソラレン（Solomon et al., Proc. Nat'l. Acad. Sci. USA 82:6470-6474, 1985; Solomon et al., Cell 53:937-947, 1988）が挙げられる。例えば、クロスリンクは、DNA分子およびクロマチンタンパク質を含む混合物に2%ホルムアルデヒドを添加することによって行うことができる。DNAをクロスリンクするために使用することができる薬剤の他の例としては、限定されないが、UV光、マイトマイシンC、ナイトロジェンマスタード、メルファラン、1,3-ブタジエンジエポキシド、シスジアンミンジクロロ白金（II）、およびシクロホスファミドが挙げられる。適切には、クロスリンク剤は、比較的短い距離（例えば、約2 Å）をクロスリンクするクロスリンクを形成し、それによって、脱クロスリンクされ得る密接な相互作用を選択する。

【0293】

いくつかの実施形態において、DNA分子は、クロスリンクの前または後に免疫沈降され得る。場合によっては、DNA分子は断片化することができる。断片は、アセチル化ヒストン、例えば、H3を特異的に認識し、それに結合する抗体などの結合パートナーと接触させることができる。このような抗体の例としては、限定されないが、Upstate Biotechnology, Lake Placid, NYから入手可能な抗アセチル化ヒストンH3が挙げられる。免疫沈降物からのポリヌクレオチドは、その後、免疫沈降物から収集され得る。クロマチンを断片化する前に、アセチル化ヒストンを隣接するポリヌクレオチド配列にクロスリンクすることができる。その後、混合物を処理して、混合物中のポリヌクレオチドを分画する。本明細書における分画技術は、デオキシリボヌクレアーゼ（DNase）酵素の使用を含む。本明細書の方法に適切なDNaseには、限定されないが、DNase I、DNase II、および小球菌ヌクレアーゼが挙げられる。得られた断片はサイズが異なり得る。得られた断片はさらに、5'末端または3'末端に一本鎖オーバーハンドを含み得る。

【0294】

いくつかの実施形態では、約145 bp～約600 bpの断片を得ることができる。代替的に、約100 bp～約2500 bp、約100 bp～約600 bp、または約600～約2500の断片を得ることができる。試料は、クロスリンクされた結合配列セグメントの配列決定のために調製することができる。場合によっては、ポリヌクレオチドの単一の短いストレッチは、例えば、分子内でクロスリンクされた2つの配列セグメントをライゲーションすることによって作製され得る。配列情報は、本明細書の他の箇所ですらに詳細に記載される任意の適切な配列決定技術、またはハイスループット配列決定法などの他の適切な方法を使用して、試料から得ることができる。例えば、ライゲーション産物を、断片の各末端から配列情報を得るペアエンド配列決定に供することができる。配列セグメントの対は、ポリヌクレオチドに沿って2つの配列セグメントを分離する直線距離にわたるハプロタイピング情報を関連付けて、得られた配列情報において表すことができる。

【0295】

H i - Cによって生成されたデータの1つの特徴は、ほとんどのリード対が、ゲノムに戻してマッピングされる場合、直鎖状に近接していることが分かることである。すなわち、ほとんどのリード対は、ゲノム中で互いに近接していることが分かる。得られたデータセットにおいて、染色体内接触の確率は、染色体が別個の区域を占有する場合に予想されるように、平均して染色体間接触の確率よりもはるかに高い。さらに、相互作用の確率は直線距離と共に急速に減衰するが、同じ染色体上で>200Mbだけ離れた遺伝子座でさえ、異なる染色体上の遺伝子座よりも相互作用する可能性が高い。ロングレンジの染色体内、および特に染色体間の接触の検出において、ショートレンジおよびミドルレンジの染色体内接触のこの「バックグラウンド」は、H i - C分析を使用して取り除くべきバックグラウンドノイズである。

10

【0296】

特に、真核生物におけるH i - C実験は、種特異的および細胞型特異的なクロマチン相互作用に加えて、2つのカノニカル相互作用パターンを示している。1つのパターンである距離依存性減衰 (distance-dependent decay) (DDD) は、ゲノム距離に応じた相互作用頻度の減衰の一般的な傾向である。第2のパターンであるシーストランス比 (CTR) は、数十メガベースの配列によって隔てられている場合でさえも、異なる染色体上の遺伝子座と比較して、有意に高い、同じ染色体上に位置する遺伝子座間の相互作用頻度である。これらのパターンは、近位の遺伝子座がランダムに相互作用するより高い確率を有する一般的なポリマー動力学、ならびに染色体領域の形成、ほとんど混合せずに核内で別個の体積を占有する傾向がある間期染色体の現象などの特定の核組織化特徴を反映し得る。これらの2つのパターンの正確な詳細は、種、細胞型、および細胞条件の間で異なり得るが、それらは遍在し、顕著である。これらのパターンは非常に強く、一貫性があるため、実験の質を評価するために使用され、通常、詳細な相互作用を明らかにするためにデータから正規化される。しかしながら、本明細書に開示される方法では、ゲノムアセンブリは、ゲノムの三次元構造を利用することができる。カノニカルH i - C相互作用パターンを、特異的なループ相互作用、すなわち、それらの遍在性、強度、および一貫性の分析のための障害にする特徴は、コンティグのゲノム位置を推定するための強力なツールとして使用することができる。

20

【0297】

特定の実施において、染色体内リード対間の物理的距離の調査は、ゲノムアセンブリに関するデータのいくつかの有用な特徴を示す。第1に、より短い距離の相互作用は、より長い距離の相互作用よりも一般的である。すなわち、リード対の各リードは、遠く離れた領域よりも実際のゲノムにおいて近い領域と結合する可能性が高い。第2に、中間距離およびロングレンジの相互作用の長いテールがある。すなわち、リード対は、キロベース (kB) またはさらにはメガベース (Mb) の距離で染色体内配置に関する情報を保持する。例えば、リード対は、約10kB、約50kB、約100kB、約200kB、約500kB、約1Mb、約2Mb、約5Mb、約10Mb、または約100Mbを超えるスパンにわたる配列情報を提供することができる。データのこれらの特徴は単に、同じ染色体上の近くにあるゲノムの領域が物理的により近接している可能性が高いことを示し、これは、それらがDNA骨格を介して互いに化学的に結合されるため、予想される結果である。H i - Cによって生成されたものなどのゲノムワイドクロマチン相互作用データセットは、染色体全体に沿った配列のグループ化および線形組織化に関するロングレンジ情報を提供すると推測された。

30

40

【0298】

H i - Cの実験方法は簡単で比較的低コストであるが、ゲノムアセンブリおよびハプロタイピングのための現在のプロトコルは、特に特定のヒト患者試料からの3~500万個の細胞、つまり、得ることが実現可能でないかもしれないかなり大量の材料を必要とする。対照的に、本明細書に開示される方法は、有意に少ない細胞由来の材料を用いて、遺伝子型アセンブリ、ハプロタイプフェージング、およびメタゲノミクスについての正確かつ

50

予測的な結果を可能にする方法を含む。例えば、約 0.1 μg 、約 0.2 μg 、約 0.3 μg 、約 0.4 μg 、約 0.5 μg 、約 0.6 μg 、約 0.7 μg 、約 0.8 μg 、約 0.9 μg 、約 1.0 μg 、約 1.2 μg 、約 1.4 μg 、約 1.6 μg 、約 1.8 μg 、約 2.0 μg 、約 2.5 μg 、約 3.0 μg 、約 3.5 μg 、約 4.0 μg 、約 4.5 μg 、約 5.0 μg 、約 6.0 μg 、約 7.0 μg 、約 8.0 μg 、約 9.0 μg 、約 10 μg 、約 15 μg 、約 20 μg 、約 30 μg 、約 40 μg 、約 50 μg 、約 60 μg 、約 70 μg 、約 80 μg 、約 90 μg 、約 100 μg 、約 150 μg 、約 200 μg 、約 300 μg 、約 400 μg 、約 500 μg 、約 600 μg 、約 700 μg 、約 800 μg 、約 900 μg 、約 1000 μg 、約 1200 μg 、約 1400 μg 、約 1600 μg 、約 1800 μg 、約 2000 μg 、約 2200 μg 、約 2400 μg 、約 2600 μg 、約 2800 μg 、約 3000 μg 、約 3200 μg 、約 3400 μg 、約 3600 μg 、約 3800 μg 、約 4000 μg 、約 4200 μg 、約 4400 μg 、約 4600 μg 、約 4800 μg 、約 5000 μg 、約 5200 μg 、約 5400 μg 、約 5600 μg 、約 5800 μg 、約 6000 μg 、約 6200 μg 、約 6400 μg 、約 6600 μg 、約 6800 μg 、約 7000 μg 、約 7200 μg 、約 7400 μg 、約 7600 μg 、約 7800 μg 、約 8000 μg 、約 8200 μg 、約 8400 μg 、約 8600 μg 、約 8800 μg 、約 9000 μg 、約 9200 μg 、約 9400 μg 、約 9600 μg 、約 9800 μg 、または約 10,000 μg 未満の DNA が、本明細書に開示される方法と共に使用することができる。いくつかの例において、本明細書に開示される方法の中で使用される DNA は、約 3,000,000、約 2,500,000、約 2,000,000、約 1,500,000、約 1,000,000、約 500,000、約 100,000、約 50,000、約 10,000、約 5,000、約 1,000、約 500、または約 100 未満の細胞から抽出することができる。

10

20

30

40

50

【0299】

普遍的に、Hi-C ベースの技術などの染色体の物理的なレイアウトをプローブするための手順は、培養細胞または初代組織から単離されたクロマチンなどの、細胞／生物内に形成されるクロマチンを利用する。本開示は、細胞／生物から単離されたクロマチンだけでなく、再構成クロマチンを用いたそのような技術の使用を提供する。再構成クロマチンは、様々な特徴にわたって細胞／生物内に形成されたクロマチンとは区別される。第 1 に、多くの試料について、ネイキッド DNA 試料の収集は、体液を集めること、頬または直腸の領域を拭き取ること、上皮試料を得ることなど、様々な非侵襲的な方法から侵襲的な方法までを使用することによって、達成することができる。第 2 に、クロマチンの再構成は、ゲノムアセンブリおよびハプロタイプフェージングのためのアーチファクトを生成する染色体間のおよび他のロングレンジの相互作用の形成を実質的に妨げる。場合によっては、試料は、約 20、15、12、11、10、9、8、7、6、5、4、3、2、1、0.5、0.4、0.3、0.2、0.1 % 未満、またはそれ以下の、本開示の方法および組成物にかかる染色体間または分子間のクロスリンキングを有してもよい。いくつかの例において、試料は、約 5 % 未満の染色体間または分子間のクロスリンキングを有してもよい。いくつかの例において、試料は、約 3 % 未満の染色体間または分子間のクロスリンキングを有してもよい。さらなる例において、試料は、約 1 % 未満の染色体間または分子間のクロスリンキングを有してもよい。第 3 に、クロスリンキングが可能な部位の頻度、したがって、ポリヌクレオチド内の分子内クロスリンキングの頻度は、調整することができる。例えば、DNA 対ヒストンの比は変動する場合があります、そうすることでヌクレオソーム密度を望ましい値に調整することができる。場合によっては、ヌクレオソーム密度は、生理的レベルより下に減らされる。したがって、クロスリンクの分布は、よりロングレンジの相互作用を支持するように変更させることができる。いくつかの実施形態において、クロスリンク密度が変動するサブ試料は、ショートレンジおよびロングレンジの両方の会合をカバーするように調製され得る。例えば、クロスリンキング条件は、クロスリンクの少なくとも約 1 %、約 2 %、約 3 %、約 4 %、約 5 %、約 6 %、約 7 %、約 8 %、約 9 %、約 10 %、約 11 %、約 12 %、約 13 %、約 14 %、約 15 %、約 16 %、約 17

%、約18%、約19%、約20%、約25%、約30%、約40%、約45%、約50%、約60%、約70%、約80%、約90%、約95%、または約100%が、試料DNA分子上で少なくとも約50kb、約60kb、約70kb、約80kb、約90kb、約100kb、約110kb、約120kb、約130kb、約140kb、約150kb、約160kb、約180kb、約200kb、約250kb、約300kb、約350kb、約400kb、約450kb、または約500kb離れているDNAセグメント間に生じるように、調整することができる。

【0300】

接触マッピングおよびトポロジー

本開示の方法によって生成されたリード対を使用して、ゲノムならびにその中の染色体および核酸分子の三次元構造を分析することができる。本明細書で論じられるように、リード対中の各リードは、ゲノム中の異なる領域にマッピングされ得る。所定のリード対について、それらがマッピングするゲノム中の2つの異なる領域は、一緒にライゲーションすることができるように、互いに空間的に近接していると推測することができる。リード対における両方のリードの座標に従って試料からのリード対をプロットすることによって、試料についての接触マップを作成することができる。例示的な接触マップは図13に見ることができ、接触マップ上の各点は、そのリード対のマッピングされた位置に従ってプロットされたリード対を表している。

【0301】

試料全体にわたる接触の分析は、染色体およびゲノムの構造の分析を可能にすることができる。ゲノムの、AおよびBコンパートメント、活性および不活性のコンパートメント、染色体コンパートメント、ユークロマチンおよびヘテロクロマチン、TADサブタイプを含むトポロジー関連ドメイン(TAD)、ならびに他の構造への組織化は、キロベースまたはメガベーススケールと同程度の規模で分析することができる。接触マップの分析は、再編成、転座、コピー数多型、逆位、欠失、および挿入などの構造的バリエーションなどのゲノム特徴の検出を可能にすることができる。

【0302】

本開示の方法は、約1bp、2bp、3bp、4bp、5bp、6bp、7bp、8bp、9bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1000bp、2000bp、3000bp、4000bp、5000bp、6000bp、7000bp、8000bp、9000bp、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、または100kb以下の解像度でタンパク質結合、構造的変化、またはゲノムの接触相互作用の位置を提供することができる。場合によっては、タンパク質結合部位、タンパク質フットプリント、接触相互作用、または他の特徴は、1000bp以内、900bp以内、800bp以内、700bp以内、600bp以内、500bp以内、400bp以内、300bp以内、200bp以内、190bp以内、180bp以内、170bp以内、160bp以内、150bp以内、140bp以内、130bp以内、120bp以内、110bp以内、100bp以内、90bp以内、80bp以内、70bp以内、60bp以内、50bp以内、40bp以内、30bp以内、20bp以内、10bp以内、9bp以内、8bp以内、7bp以内、6bp以内、5bp以内、4bp以内、3bp以内、2bp以内、または1bp以内にマッピングすることができる。一例において、本開示の方法は、ゲノム上で互いに10,000bp、5,000bp、2,000bp、または1,000bp以内にある部位(例えば、CTCF部位などのタンパク質結合部位)の解像度を可能にすることができる。場合によっては、改善された解像度またはマッピングは、保護されていない核酸(例えば、結合タンパク質のフットプリント内にはない核酸)を分解するMNaseまたは他のエンドヌクレアーゼの使用によって達成され得、それによって、保護された領域(例えば、タンパク質フットプリント)の縁で生じる近接ライゲーション事象をもたらす。

10

20

30

40

50

【0303】

コンティグマッピング

様々な実施形態において、本開示は、複数のリード対の複数のコンティグへのマッピングを可能にする様々な方法を提供する。リードをコンティグ配列にマッピングするためのいくつかの公的に入手可能なコンピュータプログラムがある。これらのリードマッピングプログラムデータはさらに、特定のリードマッピングがゲノム内でどのように固有であるかを記述するデータを提供する。固有にマッピングされるリードの集団から、コンティグ内で高い信頼度で、各リード対におけるリード間の距離の分布を推測することができる。リードが異なるコンティグに信頼できる形でマッピングされるリード対について、このマッピングデータは、問題の2つのコンティグ間の結合を意味する。これは、上述の分析から学習された距離の分布に比例する2つのコンティグ間の距離を意味する。したがって、そのリードが異なるコンティグにマッピングされる各リード対は、正しいアセンブリにおけるこれらの2つのコンティグ間の結合を意味する。すべてのそのようなマッピングされたリード対から推論される結合は、各コンティグが行および列の両方によって表される隣接行列に要約することができる。コンティグを結合するリード対は、対応する行および列において非ゼロ値としてマークされ、リード対におけるリードがマッピングされたコンティグを示す。リード対のほとんどはコンティグ内にマッピングされ、そこからリード対間の距離の分布を学習することができ、そこから異なるコンティグにマッピングされるリード対を使用してコンティグの隣接行列を構築することができる。

10

【0304】

様々な実施形態において、本開示は、リード対データからのリードマッピングデータを使用してコンティグの隣接行列を構築する工程を含む方法を提供する。いくつかの実施形態では、隣接行列は、ロングレンジの相互作用に対するショートレンジの相互作用の傾向を組み込むリード対に対する重み付けスキームを使用する。短い距離に及ぶリード対は概して、長い距離に及ぶリード対よりも一般的である。特定の距離の確率を記述する関数は、この分布を学習するために単一のコンティグにマッピングされるリード対データを使用して適合させることができる。したがって、異なるコンティグにマッピングするリード対の1つの重要な特徴は、それらがマッピングするコンティグ上の位置である。両方ともコンティグの一方の末端付近にマッピングするリード対の場合、これらのコンティグ間の推測される距離は短くなり得、したがって、結合したリード間の距離は小さい。リード対間のより短い距離は長い距離よりも一般的であるので、この構成は、これらの2つのコンティグが、コンティグの端から遠くにマッピングするリードよりも隣接しているというより強力な証拠を提供する。したがって、隣接行列における結合は、コンティグの端までのリードの距離によってさらに重み付けされる。さらなる実施形態では、隣接行列をさらに再スケールして、ゲノムの無差別領域を表すいくつかのコンティグ上の多数の接触点の重みを下げることができる。ゲノムのこれらの領域は、それらに対して高い割合のリードマッピングを有することによって同定可能であり、先験的に、アセンブリを誤認する可能性のある偽のリードマッピングを含む可能性が高い。またさらなる実施形態において、このスケールリングは、転写抑制因子CTCF、内分泌受容体、コヒーシン、または共有結合的に修飾されたヒストンなどの、クロマチンの足場相互作用を調節する1つ以上の薬剤についての1つ以上の保存された結合部位を探索することによって指向され得る。

20

30

40

【0305】

いくつかの実施形態では、本開示は、ゲノムに対する順序および／または配向を表すコンティグを通る経路を決定するために、隣接行列を分析する工程を含む、本明細書に開示される1つ以上の方法を提供する。他の実施形態では、コンティグを通る経路は、各コンティグを正確に1回通過するように選択することができる。さらなる実施形態では、コンティグを通る経路は、隣接行列を通る経路が通過したエッジの重みの合計を最大化するように選択される。このようにして、正しいアセンブリのために最も可能性の高いコンティグ結合が提案される。またさらなる実施形態では、コンティグを通る経路は、各コンティグを正確に1回通過し、隣接行列のエッジの重み付けが最大化されるように選択すること

50

ができる。

【0306】

ハプロタイプフェージング

二倍体ゲノムでは、どの対立遺伝子バリエーションが同じ染色体上に結合されているかを知ることがしばしば重要である。これはハプロタイプフェージングとして知られている。ハイスループット配列データからのショートリードは、どの対立遺伝子バリエーションが結合されているかを直接観察することを可能にすることはめったにない。ハプロタイプフェージングのコンピュータによる推論は、本開示は、リード対上の対立遺伝子バリエーションを使用して、どの対立遺伝子バリエーションが結合されているかを決定することを可能にする1つ以上の方法を提供する。場合によっては、本開示の方法によるフェージングは、補完なしで行われる。

10

【0307】

様々な実施形態において、本開示の方法および組成物は、複数の対立遺伝子バリエーションに関して、二倍体または多倍体のゲノムのハプロタイプフェージングを可能にする。したがって、本明細書に記載される方法は、リード対からのバリエーション情報および／またはそれを使用するアセンブルされたコンティグに基づいて、結合された対立遺伝子バリエーションの決定を提供することができる。対立遺伝子バリエーションの例としては、限定されないが、1000ゲノム、UK10K、HapMap、およびヒト間の遺伝的変異を発見するための他のプロジェクトから知られているものが挙げられる。場合によっては、例えば、シャルコー・マリー・トゥース神経障害 (Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. N. Engl. J. Med. 362:1181-91, 2010) を引き起こす SH3TC2 の両方のコピーにおける非連鎖の不活性化突然変異、および高コレステロール血症9 (Rios J, Stein E, Shendure J, et al. Hum. Mol. Genet. 19:4313-18, 2010) を引き起こす ABCG5 の両方のコピーにおける非連鎖の不活性化突然変異の発見によって実証されるように、ハプロタイプフェージングデータを有することによって、特定の遺伝子への疾患の関連がより容易に明らかにされ得る。

20

【0308】

ヒトは、1,000において平均1部位でヘテロ接合性である。場合によっては、ハイスループット配列決定法を使用するデータの単一レーンは、少なくとも約150,000,000個のリード対を生成することができる。リード対は約100塩基対長であり得る。これらのパラメーターから、ヒト試料からの全リードの10分の1が、ヘテロ接合部位をカバーすると推定される。したがって、ヒト試料由来の全リード対の平均100分の1は、ヘテロ接合部位の対をカバーすると推定される。したがって、約1,500,000リード対 (150,000,000の百分の1) は、単一レーンを使用してフェージングデータを提供する。ヒトゲノム中の約30億塩基、および1000中の1つがヘテロ接合性である場合、平均ヒトゲノム中に約300万のヘテロ接合性部位が存在する。ヘテロ接合性部位の対を表す約1,500,000個のリード対では、ハイスループット配列法の単一レーンを使用してフェージングされる各ヘテロ接合性部位の平均カバレッジは、典型的なハイスループット配列決定機を使用して、約 (1倍 (1X)) である。したがって、二倍体ヒトゲノムは、本明細書に開示される方法を使用して調製される試料由来の配列バリエーションに関連するハイスループット配列データの1つのレーンと、確実に完全なフェージングすることができる。いくつかの例では、データのレーンは、DNA配列リードデータのセットであり得る。さらなる例において、データのレーンは、ハイスループット配列決定機の1回のランからのDNA配列リードデータのセットであり得る。

30

40

【0309】

ヒトゲノムは染色体の2つの相同セットからなるため、個体の真の遺伝子構成を理解することは、遺伝物質の母体および父体のコピーまたはハプロタイプの描写を必要とする。個体においてハプロタイプを得ることは、いくつかの方法で有用である。第1に、ハプロタイプは、臓器移植におけるドナー-宿主適合の結果の予測において臨床的に有用であり

50

、疾患関連性を検出するための手段としてますます使用されている。第2に、複合ヘテロ接合性を示す遺伝子において、ハプロタイプは、2つの有害なバリエーションが同じ対立遺伝子上に位置するかどうかに関する情報を提供し、これらのバリエーションの遺伝が有害であるかどうかの予測に大きく影響する。第3に、個体群からのハプロタイプは、人種の集団構造および進化歴に関する情報を提供した。最後に、最近記載された遺伝子発現における広範な対立遺伝子不均衡は、対立遺伝子間の遺伝的または後成的な差異が発現の定量的差異に寄与し得ることを示唆している。ハプロタイプ構造の理解は、対立遺伝子不均衡に寄与するバリエーションの機構を描写するであろう。

【0310】

特定の実施形態では、本明細書に開示される方法は、ロングレンジの結合およびフェーシングのために必要なゲノムの遠位領域間の会合を固定および捕捉するためのインビトロ技術を含む。場合によっては、方法は、非常にゲノム的に遠位のリード対を送達するために、XLRPライブラリーを構築および配列決定する工程を含む。場合によっては、相互作用は主として単一のDNA断片内のランダムな会合から生じる。いくつかの例では、DNA分子において互いに近接する配列セグメントがより頻繁にかつより高い確率で相互作用するが、分子の遠位部分間の相互作用はより低頻度であるため、配列セグメント間のゲノム距離を推測することができる。したがって、2つの遺伝子座を結合する対の数と入力DNA上でのそれらの近接性との間には系統的な関係がある。本開示は、抽出において最大のDNA断片に及ぶことができるリード対を生成することができる。このライブラリーのための入力DNAは、150kbpの最大長を有し、これは、配列決定データから観察される最も長い意味のあるリード対である。これは、より大きな入力DNA断片が提供される場合、本方法がなおよりゲノム的に遠位の遺伝子座を結合することができることを示唆している。本方法によって生成されたデータのタイプを取り扱うように特に適合された改良型のアセンブリソフトウェアツールを適用することによって、完全なゲノムアセンブリが可能であり得る。

【0311】

極めて高いフェーシング精度は、本開示の方法および組成物を使用して生成されたデータによって達成することができる。以前の方法と比較して、本明細書に記載される方法は、より高い割合のバリエーションをフェーシングすることができる。フェーシングは、高いレベルの精度を維持しながら達成することができる。本明細書の技術は、約70%、80%、90%、95%、96%、97%、98%、99%、99.9%、99.99%、または99.999%を超える精度でのフェーシングを可能にすることができる。本明細書の技術は、約500×配列決定深さ、450×配列決定深さ、400×配列決定深さ、350×配列決定深さ、300×配列決定深さ、250×配列決定深さ、200×配列決定深さ、150×配列決定深さ、100×配列決定深さ、または50×配列決定深さ未満の正確なフェーシングを可能にすることができる。このフェーシング情報は、より長い距離、例えば、約200kbp、約300kbp、約400kbp、約500kbp、約600kbp、約700kbp、約800kbp、約900kbp、約1Mbp、約2Mbp、約3Mbp、約4Mbp、約5Mbp、または約10Mbp超に拡張することができる。いくつかの実施形態において、ヒト試料についてのヘテロ接合性SNPの90%超は、約2.5億未満のリードまたはリード対を使用して、例えば、Illumina HiSeqデータの1レーンのみを使用することによって、99%を超える精度でフェーシングすることができる。他の場合において、ヒト試料についてのヘテロ接合性SNPの約40%、50%、60%、70%、80%、90%、95%、または99%超が、約2.5億未満または約5億未満のリードまたはリード対を使用して、例えば、Illumina HiSeqデータの1または2レーンのみを使用することによって、約70%、80%、90%、95%、96%、97%、98%、99%、99.9%、99.99%、または99.999%を超える精度でフェーシングすることができる。例えば、ヒト試料についてのヘテロ接合性SNPの95%または99%超は、約2.5億未満または約5億未満のリードを使用して、約95%または99%を超える精度でフェーシングすることができる。

さらなる場合において、リード長を約 200 bp、250 bp、300 bp、350 bp、400 bp、450 bp、500 bp、600 bp、800 bp、1000 bp、1500 bp、2 kb、3 kb、4 kb、5 kb、10 kb、20 kb、50 kb、または 100 kb に増加させることによって、さらなるバリエーションを捕捉することができる。

【0312】

本開示の他の実施形態では、XLRPライブラリーからのデータを使用して、ロングレンジのリード対のフェージング能力を確認することができる。これらの結果の精度は、以前に利用可能な最良の技術と同等であるが、有意に長い距離までさらに拡大する。特定の配列決定方法のための現在の試料調製プロトコルは、フェージングのための標的部位のリード長、例えば、150 bp 内に位置するバリエーションを認識する。一例では、アセンブリのベンチマーク試料である NA12878 について構築された XLRP ライブラリーから、存在する 1,703,909 のヘテロ接合性 SNP の 44 % を、99 % を超える精度でフェージングした。場合によっては、この割合は、酵素の賢明な選択または消化条件を用いて、ほぼすべての可変部位に拡大することができる。

【0313】

ハプロタイプフェージングは、ヒト白血球抗原 (HLA) 領域 (例えば、クラス I HLA-A、B、および C、クラス II HLA-DRB1/3/4/5、HLA-DQA1、HLA-DQB1、HLA-DPA1、HLA-DPB1) のフェージングを含むことができる。ゲノムの HLA 領域は密集した多型であり、標準的な配列決定アプローチでは配列決定またはフェージングすることが困難であり得る。本開示の技法は、ゲノムの HLA 領域の改善された配列決定およびフェージング精度を提供することができる。本開示の技法を用いて、ゲノムの HLA 領域は、より大きな領域 (例えば、染色体アーム、染色体、全ゲノム) またはそれ自体 (例えば、ハイブリッド捕捉などの標的濃縮による) のフェージングの一部として正確にフェージングすることができる。一例では、HLA 領域自体を、約 300 倍 (x) の配列決定深さで正確にフェージングした。これらの技法は、ロングレンジ PCR などの HLA 分析のための従来のアプローチに勝る利点を提供することができる。例えば、ロングレンジ PCR は、複雑なプロトコルおよび多くの別個の反応を含むことができる。本明細書でさらに論じられるように、例えば、試料識別バーコードを架橋オリゴヌクレオチドまたは他の場所に含め、バーコードに基づいて配列情報を逆多重化することによって、配列決定分析のために試料を多重化することができる。例えば、複数の試料を近接ライゲーションに供し、試料同定バーコード (例えば、架橋オリゴヌクレオチド中) でバーコード化し、HLA 領域を標的化し (例えば、ハイブリッド捕捉によって)、多重配列決定を行い、複数の試料についての HLA 領域のフェージングを可能にする。場合によっては、HLA 領域のフェージングは、補完なしで行われる。

【0314】

ハプロタイプフェージングは、キラー細胞免疫グロブリン様受容体 (KIR) 領域のフェージングを含むことができる。ゲノムの KIR 領域は、トランスポゾン媒介性組換えのために高度に相同かつ構造的に動的であり、標準的な配列決定アプローチを用いて配列決定またはフェージングすることは困難であり得る。本開示の技法は、ゲノムの KIR 領域の改善された配列決定およびフェージング精度を提供することができる。本開示の技法を用いて、ゲノムの KIR 領域は、より大きな領域 (例えば、染色体アーム、染色体、全ゲノム) またはそれ自体 (例えば、ハイブリッド捕捉などの標的濃縮による) のフェージングの一部として正確にフェージングすることができる。これらの技法は、ロングレンジ PCR などの HLA 分析のための従来のアプローチに勝る利点を提供することができる。例えば、ロングレンジ PCR は、複雑なプロトコルおよび多くの別個の反応を含むことができる。本明細書でさらに論じられるように、例えば、試料識別バーコードを架橋オリゴヌクレオチドまたは他の場所に含め、バーコードに基づいて配列情報を逆多重化することによって、配列決定分析のために試料を多重化することができる。例えば、複数の試料を近接ライゲーションに供し、試料同定バーコード (例えば、架橋オリゴヌクレオチド中) で

バーコード化し、K I R 領域を標的化し（例えば、ハイブリッド捕捉によって）、多重配列決定を行い、複数の試料についての K I R 領域のフェージングを可能にする。少なくとも約 8、9、10、11、12、13、14、15、16、17、またはそれ以上の遺伝子および／または偽遺伝子をフェージングすることができる。場合によっては、K I R 領域のフェージングは、補完なしで行われる。

【0315】

メタゲノミクス分析

いくつかの実施形態では、本明細書に記載される組成物および方法は、メタゲノム、例えば、ヒトの腸で見られるものの調査を可能にする。したがって、所定の生態学的環境に生息するいくつかまたはすべての生物の部分的または完全なゲノム配列を調査することができる。例としては、すべての腸内微生物、皮膚の特定の領域で見られる微生物、および毒性廃棄物部位に生存する微生物のランダムな配列決定が挙げられる。これらの環境における微生物集団の組成は、本明細書に記載される組成物および方法、ならびにそれらのそれぞれのゲノムによってコードされる相互に関連する生化学の態様を使用して、決定することができる。本明細書に記載される方法は、複雑な生物学的環境、例えば、2、3、4、5、6、7、8、9、10、12、15、20、25、30、40、50、60、70、80、90、100、125、150、175、200、250、300、400、500、600、700、800、900、1000、5000、10000 またはそれ以上を超える生物および／または生物のバリエーションを含む生物学的環境からのメタゲノム研究を可能にすることができる。

【0316】

癌ゲノム配列決定に必要とされる高い精度は、本明細書に記載される方法およびシステムを使用して達成することができる。不正確な参照ゲノムは、癌ゲノムを配列決定する際に塩基呼び出しを困難なものにしかねない。不均一な試料および小さな出発物質、例えば、生検により得られた試料はさらなる困難をもたらす。さらに、大規模な構造的なバリエーションおよび／またはヘテロ接合性の損失の検出はしばしば、癌ゲノム配列決定の他、体細胞バリエーションと塩基呼び出しのエラーとを区別する能力にも、必要不可欠である。

【0317】

改善された配列決定精度

本明細書に記載されるシステムおよび方法は、2、3、4、5、6、7、8、9、10、12、15、20、またはそれ以上の様々なゲノムを含む複雑な試料から正確な長い配列を生成し得る。通常の、良性の、および／または腫瘍由来の混合試料は任意選択で、通常の対照を必要とすることなく、解析されてもよい。いくつかの実施形態において、100 ng または数百ほどのゲノム当量の小さな出発試料は、正確な長い配列を生成するために利用される。本明細書に記載されるシステムおよび方法は、大規模な構造バリエーションおよび再編成の検出が可能となる場合があり、フェージングされたバリエーションコールは、約 1 k b p、約 2 k b p、約 5 k b p、約 10 k b p、約 20 k b p、約 50 k b p、約 100 k b p、約 200 k b p、約 500 k b p、約 1 M b p、約 2 M b p、約 5 M b p、約 10 M b p、約 20 M b p、約 50 M b p、または約 100 M b p、またはそれ以上のヌクレオチドにまたがる長い配列にわたって得られ得る。例えば、フェージングされたバリエーションコールは、約 1 M b p または約 2 M b p にまたがる長い配列にわたって得られる。

【0318】

本明細書に記載される方法およびシステムを使用して決定されるハプロタイプは、コンピュータリソース、例えば、クラウドシステムなどのネットワーク上のコンピュータリソースに割り当てられてもよい。短いバリエーションコールは、必要であれば、コンピュータリソースに保管される関連情報を使用して修正することができる。構造バリエーションは、短いバリエーションコールからの組み合わせられた情報、およびコンピュータリソースに保管された情報に基づいて検出することができる。ゲノムの問題となる部分、例えば、セグメント重複、構造多型 (s t r u c t u r a l v a r i a t i o n) の傾向がある領域、高度に

可変的でありかつ医学的に関連するMHC領域、セントロメアおよびテロメアの領域、ならびに、限定されないが、反復領域、低い配列精度、高いバリエーション率、A L U反復、セグメント重複を持つものを含む他の異質染色質領域、または当該技術分野での他の関連する問題となる部分を、精度の向上ために再度組み立てることができる。

【0319】

試料の種類は、局所的に、またはクラウドなどのネットワーク接続されたコンピュータリソースにおいて、配列情報に割り当てることができる。情報のソースが分かっている場合、例えば、情報のソースが癌または正常な組織由来である場合、このソースは試料の種類の一部として試料に割り当てることができる。他の試料の種類の場合は通常、限定されないが、組織の種類、試料収集方法、感染の存在、感染の種類、処理方法、試料のサイズなどを含む。癌ゲノムとの比較における通常のゲノムのなどの、完全または部分的な比較ゲノム配列が利用可能である場合、試料データと比較ゲノム配列との差を判定することができる、任意選択で出力することができる。

【0320】

臨床用途

本開示の方法は、選択的なゲノムの関心領域の他に、選択的な関心領域と相互作用し得るゲノム領域の遺伝子情報の分析に使用され得る。本明細書に開示されるような増幅方法は、限定されないが、米国特許第6,449,562号、第6,287,766号、第7,361,468号、第7,414,117号、第6,225,109号、および第6,110,709号でみられるものなど、遺伝子分析のためのデバイス、キット、および方法に使用することができる。場合によっては、本開示の増幅方法は、多型の存在の有無を判定するためのDNAハイブリダイゼーション研究のために標的核酸を増幅するために使用することができる。多型または対立遺伝子は、遺伝病などの疾患または疾病に関連付けることができる。他の場合において、多型は、疾患または疾病に対する感受性に関連付けることができ、例えば、多型は、中毒、退行性および年齢に関連する疾病、癌などに関連付けられる。他の場合において、多型は、冠動脈の健康の増大といった有用な特色、HIVもしくはマラリアなどの疾患に対する抵抗性、または骨粗鬆症、アルツハイマー病、もしくは痴呆などの成人病に対する抵抗性に関連付けることができる。

【0321】

本開示の組成物および方法は、診断、予後、治療、患者の層別化、薬物の開発、処置の選択、およびスクリーニングの目的のために使用することができる。本開示は、本開示の方法を使用して多くの様々な標的分子が単一の生体分子試料から一度に分析することができるという利点を提供する。これにより、例えば、様々な診断試験を1つの試料上で行うことが可能となる。

【0322】

本開示の組成物および方法はゲノミクスにおいて使用され得る。本明細書に記載される方法は、この用途に非常に望ましい答えを迅速に示すことができる。本明細書に記載される方法および組成物は、診断または予後のためにかつ健康と疾患の指標として使用され得るバイオマーカーを見つけ出すプロセスで使用され得る。本明細書に記載される方法および組成物は、薬物をスクリーニングするために、例えば、薬物の開発、処置の選択、処置の有効性の判定、および／または医薬開発のための標的の識別のために、使用され得る。薬物に関するスクリーニングアッセイ時に遺伝子発現を試験する能力は、タンパク質が身体において最終的な遺伝子産物であるため、非常に重要である。いくつかの実施形態において、本明細書に記載される方法および組成物は、実行されている特定のスクリーニングに関する最も多くの情報を提供するタンパク質および遺伝子発現の両方を同時に測定する。

【0323】

本開示の組成物および方法は、遺伝子発現分析に使用され得る。本明細書に記載される方法は、ヌクレオチド配列を区別する。標的ヌクレオチド配列間の差は、例えば、単一の核酸塩基の差、核酸欠失、核酸挿入、または再編成であり得る。1を超える塩基を含むそ

のような配列の差も検出することができる。本開示のプロセスは、感染症、遺伝子疾患、および癌を検出することができる。さらに、上記プロセスは環境モニタリング、法医学、および食品科学においても有用である。核酸に対して実施され得る遺伝子分析の例としては、例えば、SNP検出、STR検出、RNA発現分析、プロモーターメチル化、遺伝子発現、ウイルス検出、ウイルスのサブタイプ分類、および薬物耐性が挙げられる。

【0324】

本方法は、罹患した細胞型が試料に存在するかどうか、疾患の段階、患者の予後、特定の処置に応答する患者の能力、または患者にとって最良の処置を判定するために、患者から得られたまたは患者に由来する生体分子試料の分析に適用され得る。本方法は、特定の疾患のためのバイオマーカーを同定するためにも適用され得る。

10

【0325】

いくつかの実施形態において、本明細書に記載される方法は疾病の診断に使用される。本明細書で使用されるように、「診断する」または疾病の「診断」という用語は、疾病を予測もしくは診断すること、疾病の素因を判定すること、疾病の処置をモニタリングすること、疾患の治療反応、もしくは疾病の予後、疾病の進行、または疾病の特定の処置に対する反応を診断することを含む。例えば、血液試料は、試料中の疾患または悪性細胞型のマーカーの存在および／または量を判定するために、本明細書に記載される方法のうちいずれかに従ってアッセイすることができ、それによって、疾患または癌を診断またはステージ分類することができる。

【0326】

20

いくつかの実施形態において、本明細書に記載される方法および組成物は、疾病の診断および予後のために使用される。

【0327】

多数の免疫学的、増殖性、および悪性の疾患と障害が、本明細書に記載される方法に特に適している。免疫疾患および障害は、アレルギー性疾患および障害、免疫機能の障害、ならびに自己免疫疾患および疾病を含む。アレルギー性疾患および障害は、限定されないが、アレルギー性鼻炎、アレルギー性結膜炎、アレルギー性喘息、アトピー性湿疹、アトピー性皮膚炎、および食物アレルギーを含む。免疫不全症は、限定されないが、重症複合免疫不全症（SCID）、好酸球増加症候群、慢性肉芽腫症、白血球接着不全症IおよびII、高IgE症候群、チェディアック・東、好中球増加症、好中球減少症、無形成症、無ガンマグロブリン血症、高IgM症候群、ディジョージ／軟口蓋帆・心臓・顔症候群、およびインターフェロン・ガンマ-T_H1経路欠損症を含む。自己免疫性および免疫調節異常の障害は、限定されないが、関節リウマチ、糖尿病、全身性エリトマトーデス、グレーブス病、グレーブス眼症、クローン病、多発性硬化症、乾癬、全身性硬化症、甲状腺腫およびリンパ腫性甲状腺腫（橋本甲状腺炎、リンパ節様甲状腺腫）、円形脱毛症、自己免疫性心筋炎、硬化性苔癬、自己免疫性ブドウ膜炎、アジソン病、萎縮性胃炎、重症筋無力症、特発性血小板減少性紫斑病、溶血性貧血、原発性胆汁性肝硬変、ウェゲナー肉芽腫症、結節性多発動脈炎、および炎症性腸疾患、同種異系移植片拒絶反応、ならびに感染性細菌または環境抗原に対するアレルギー反応による組織破壊を含む。

30

【0328】

40

本開示の方法により評価され得る増殖性疾患および障害は、限定されないが、新生児の血管腫症、二次性進行型多発性硬化症、慢性進行性骨髄変性疾患、神経線維腫症、神経節神経腫症、ケロイド形成、骨のパジェット病、（例えば、乳房または子宮の）線維嚢胞症、サルコイドーシス、ペロニーおよびデュピュイトランの繊維症、硬変、アテローム性動脈硬化、および血管再狭窄を含む。

【0329】

本開示の方法により評価され得る悪性疾患および障害は、血液悪性腫瘍および固形腫瘍の両方を含む。

【0330】

血液悪性腫瘍は特に、血液由来の細胞の変化を含むため、試料が血液試料である時に本

50

開示の方法に適している。そのような悪性腫瘍は、非ホジキンリンパ腫、ホジキンリンパ腫、非B細胞リンパ腫、および他のリンパ腫、急性または慢性の白血病、多血球血症、血小板血症、多発性骨髄腫、骨髄異形成障害、骨髄増殖性障害、骨髄線維症 (myelofibroses)、異型免疫リンパ球増殖、および形質細胞障害を含む。

【0331】

本開示の方法により評価され得る形質細胞疾患は、多発性骨髄腫、アミロイドーシス、およびヴァルデンストレームマクログロブリン血症を含む。

【0332】

固形腫瘍の例としては、限定されないが、結腸癌、乳癌、肺癌、前立腺癌、脳腫瘍、中枢神経系腫瘍、膀胱腫瘍、黒色腫、肝臓癌、骨肉腫、および他の骨癌、睾丸および卵巣の癌腫、頭頸部腫瘍、および子宮頸部の新生物が挙げられる。

【0333】

遺伝子疾患も本開示のプロセスにより検出することができる。これは、染色体および遺伝子の異常、または遺伝子疾患に関する出生前または出生後のスクリーニングにより実行され得る。検出可能な遺伝子疾患の例は、21ヒドロキシラーゼ欠損、嚢胞性繊維症、脆弱X症候群、ターナー症候群、デュシェンヌ型筋ジストロフィー、ダウン症候群または他のトリソミー、心疾患、単一の遺伝子疾患、HLA分類、フェニルケトン尿症、鎌状赤血球貧血、テイ・ザックス病、サラセミア、クラインフェルター症候群、ハンチントン病、自己免疫疾患、リビドーシス、肥満欠損症 (obesity defect)、血友病、先天性代謝異常症、および糖尿病を含む。

【0334】

本開示の方法は、限定されないが、遺伝子融合、構造バリエーション、再編成、およびトポロジーの変化、例えば、TAD境界の欠失または変化、TADサブタイプの変化、コンパートメントの変化、クロマチン型の変化、および修飾状態の変化、例えば、メチル化状態 (例えば、CpGメチル化、H3K4me3、H3K27me3、または他のヒストンメチル化) を含む遺伝的疾患に関連する遺伝的特徴またはゲノム特徴を検出するために使用され得る。

【0335】

本明細書に記載される方法は、試料中の細菌またはウイルスそれぞれのマーカーの存在および／または量を判定することにより、病原体感染、例えば、細胞内細菌およびウイルスによる感染を診断するために使用され得る。

【0336】

多種多様な感染症が本開示のプロセスにより検出することができる。感染症は、細菌、ウイルス、寄生生物、および真菌の感染因子により引き起こされ得る。薬物に対する様々な感染因子の抵抗性も、本開示を使用して判定され得る。

【0337】

本開示により検出され得る細菌感染因子は、エシェリキア・コリ、サルモネラ、シゲラ、クレブシエラ、シュドモナス、リステリア・モノサイトゲネス、マイコバクテリウム・ツベルクローシス、マイコバクテリウム・アビウム・イントラセルラーレ、エルシニア、フランシセラ、パスツレラ、ブルセラ、クロストリジウム、ボルデテラ・ペルツシス、バクテロイデス、スタフィロコッカス・アウレウス、ストレプトコッカス・ニューモニエ、B溶血性連鎖球菌 (B-Hemolytic strep.)、コリネバクテリア、レジオネラ、ミコプラズマ、ウレアプラスマ、クラミジア、ナイセリア・ゴノレエ (Neisseria gonorrhoeae)、ナイセリア・メニンギティディス、ヘモフィルス・インフルエンザ、エンテロコッカス・フェカリス、プロテウス・ブルガリス、プロテウス・ミラビリス、ヘリコバクター・ピロリ、トレポネマ・パラジウム、ボレリア・ブルグドルフェリ、ボレリア・リカレンチス、リケッチア病原体、ノカルディア、および放線菌 (Actinomycetes) を含む。

【0338】

本開示により検出され得る真菌感染因子は、クリプトコッカス・ネオフォルマンズ、ブ

10

20

30

40

50

ラストミセスーデルマティティディス、ヒストプラスマーカプスラーツム、コクシジオイデスーイミチス、パラコクシジオイデスーブラジリエンシス、カンジダーアルビカンス、アスペルギルスーフミガーツス (*Aspergillus fumigatus*)、藻菌類 (リゾープス)、スポロトリックスーシェンキー、クロモミコーシス、およびマズラミコーシスを含む。

【0339】

本開示により検出され得るウイルス感染因子は、ヒト免疫不全ウイルス、ヒトT細胞リンパ性細胞栄養性ウイルス (*human T-cell lymphocytotropic virus*)、肝炎ウイルス (例えば、B型肝炎ウイルスおよびC型肝炎ウイルス)、エプスタインバーウイルス、サイトメガロウイルス、ヒトパピローマウイルス、オルソミクソウイルス、パラミクソウイルス、アデノウイルス、コロナウイルス、ラブドウイルス、ポリオウイルス、トogaウイルス、ブンヤウイルス、アレナウイルス、風疹ウイルス、およびレオウイルスを含む。

【0340】

本開示により検出され得る寄生生物因子は、熱帯熱マラリア原虫、四日熱マラリア原虫、三日熱マラリア原虫、卵型マラリア原虫、回旋糸状虫 (*Onchocerca volvulus*)、リーシュマニア、トリパノゾーマ種、住血吸虫種、エントアメーバーヒストリチカ、クリプトスポリジウム (*Cryptosporidium*)、ジアルジア種、トリコモナス (*Trichomonas*) 種、大腸バランチジウム (*Balantidium Coli*)、バンクロフト糸状虫、トキソプラズマ種、蟯虫、回虫、鞭虫、メジナ虫 (*Dracunculus medinensis*)、吸虫、広節裂頭条虫、テニア種、ニューモシスチスーカリニ、およびアメリカ鉤虫 (*Necator americanis*) を含む。

【0341】

本開示は、感染因子による薬物抵抗性の検出にも有用である。例えば、バイコマイシン耐性エンテロコッカス・フェシウム、メチシリン耐性スタフィロコッカスーアウレウス、ペニシリン耐性ストレプトコッカスーニューモニエ、多剤耐性マイコバクテリウムーツベルクローシス、および耐AZT性ヒト免疫不全ウイルスはすべて、本開示で同定され得る。

【0342】

したがって、本開示の組成物および方法を使用して検出された標的分子は、患者のマーカー (癌マーカーなど)、または細菌もしくはウイルスマーカーなどの異物による感染のマーカーのいずれかであり得る。

【0343】

本開示の組成物および方法は、標的分子を同定および／または定量化するために使用され得、標的分子の存在量は、生物学的状態または疾患条件、例えば、疾患状態の結果としてアップレギュレートまたはダウンレギュレートされる血液マーカーを示す。

【0344】

いくつかの実施形態において、本開示の方法および組成物は、サイトカイン発現のために使用することができる。本明細書に記載される方法の低感受性は、例えば、癌などの疾患の状態、診断、または予後のバイオマーカーとしてのサイトカインの早期発見、および亜臨床的な状態の同定に有用である。

【0345】

本開示の方法は、限定されないが、遺伝子融合、構造バリエーション、再編成、およびトポロジーの変化、例えば、TAD境界の欠失または変化、TADサブタイプの変化、コンパートメントの変化、クロマチン型の変化、および修飾状態の変化、例えば、メチル化状態 (例えば、CpGメチル化、H3K4me3、H3K27me3、または他のヒストンメチル化) を含む癌に関連する遺伝的特徴またはゲノム特徴を検出するために使用され得る。

【0346】

試料

標的ポリヌクレオチドが由来する様々な試料は、同じ個体の複数の試料、異なる個体の試料、またはそれらの組み合わせを含み得る。いくつかの実施形態において、試料は、1人の個体の複数のポリヌクレオチドを含む。いくつかの実施形態において、試料は、2人以上の個体の複数のポリヌクレオチドを含む。個体は、標的ポリヌクレオチドが由来し得る任意の生物またはその一部であり、その非限定的な例としては、植物、動物、真菌、原生生物、モネラ界、ウイルス、ミトコンドリア、および葉緑体が挙げられる。試料ポリヌクレオチドは、細胞試料、組織試料、または、例えば、培養細胞株、生検、血液試料、もしくは細胞を含有している流体試料を含む、それらに由来する臓器試料など、対象から単離することができるものである。対象は、限定されないが、ウシ、ブタ、マウス、ラット、ニワトリ、ネコ、イヌなどの動物を含む動物であってもよく、通常はヒトなどの哺乳動物である。試料は化学合成などにより人工的に得られることもできる。いくつかの実施形態において、試料はDNAを含む。いくつかの実施形態において、試料はゲノムDNAを含む。いくつかの実施形態において、試料は、ミトコンドリアDNA、葉緑体DNA、プラスミドDNA、バクテリア人工染色体、酵母人工染色体、オリゴヌクレオチドタグ、またはそれらの組み合わせを含む。いくつかの実施形態において、試料は、限定されないが、ポリメラーゼ連鎖反応（PCR）、逆転写、およびそれらの組み合わせを含む、プライマーとDNAポリメラーゼとの適切な組み合わせを使用したプライマー伸長反応により生成されたDNAを含む。プライマー伸長反応のための鋳型がRNAである場合、逆転写の産物は相補的DNA（cDNA）と呼ばれる。プライマー伸長反応に有用なプライマーは、1つ以上の標的に特異的な配列、ランダム配列、部分的なランダム配列、およびそれらの組み合わせを含み得る。プライマー伸長反応に適した反応条件は当該技術分野で知られている。一般に、試料のポリヌクレオチドは、試料中に存在する任意のポリヌクレオチドを含み、これは標的ポリヌクレオチドを含むこともあれば、含まないこともある。

10

20

30

40

50

【0347】

いくつかの実施形態において、核酸鋳型分子（例えば、DNAまたはRNA）は、タンパク質、脂質、および非鋳型核酸などの、様々な他の成分を含有する生体試料から単離される。核酸鋳型分子は任意の細胞材料から得られ、動物、植物、細菌、真菌類、または他の細胞生物から得られ得る。本開示での使用のための生体試料は、ウイルスの粒子または調製物を含む。核酸鋳型分子は、生物から直接、または、生物から得た生体試料、例えば、血液、尿、脳脊髄液、精液、唾液、痰、便、および組織から得ることができる。任意の組織または体液の標本が、本開示で使用される核酸の供給源として使用されてもよい。核酸鋳型分子は、初代細胞培養物または細胞株などの培養細胞からも単離され得る。鋳型核酸が得られる細胞または組織は、ウイルスまたは他の細胞内病原体に感染され得る。試料は、生体標本、cDNAライブラリー、ウイルスDNA、またはゲノムDNAから抽出された総体的なRNAでもあり得る。試料は、細胞構造がない起源から単離されたDNA、例えば、冷凍装置から増幅／単離されたDNAでもあり得る。

【0348】

核酸の抽出と精製のための方法は公知である。例えば、核酸は、フェノール、フェノール／クロロホルム／イソアミルアルコール、またはTRIzolおよびTriReagentを含む同様の製剤での有機抽出により精製され得る。抽出技術の他の非限定的な例は、（1）自動核酸抽出器、例えば、Applied Biosystems（Foster City, Calif.）から入手可能なModel 341 DNA Extractorの使用を伴うまたは伴わない、例えば、フェノール／クロロホルムの有機試薬（Ausubel et al., 1993）を使用する、有機抽出とその後のエタノール沈殿、（2）固定相吸着法（米国特許第5,234,809号、Walsh et al., 1991）、および（3）典型的に「塩析」方法と呼ばれる沈澱法などの、塩で誘導された核酸沈澱法（Miller et al., (1988)）を含む。核酸の単離および／または精製の別の例は、核酸が特異的または非特異的に結合することができる磁性粒子の使用と、その後の、磁石を使用したビーズの単離と、核酸の洗浄と、ビーズから

の核酸の溶出を含む（例えば、米国特許第5,705,628号を参照）。いくつかの実施形態では、上記の単離方法の前に、試料から望ましくないタンパク質を除去するのを助けるための酵素消化工程、例えば、プロテイナーゼKまたは他の同様のプロテアーゼによる消化を行ってもよい（例えば、米国特許第7,001,724号を参照されたい）。必要に応じて、RNAse阻害剤を溶解緩衝液に添加することができる。特定の細胞または試料型について、前記プロトコルにタンパク質変性／消化の工程を加えることが望ましい場合もある。精製方法は、DNA、RNA、またはその両方を単離することを目的とし得る。抽出手順の間またはその後に、DNAとRNAの両方が一緒に単離されると、さらなる工程を利用して、一方または両方を他とは別々に精製することができる。例えば、サイズ、配列、または他の物理的もしくは化学的な特性による精製により、抽出した核酸の細画分を生成することもできる。最初の核酸単離工程に加えて、過剰なまたは不要な試薬、反応物、または生成物を除去するなどのために、本開示の方法における工程の後に、核酸の精製を実施することができる。

10

【0349】

核酸の鑄型分子は、2003年10月9日に公開された米国特許出願公開第US2002/0190663号A1に記載される通りに得ることができる。通常、核酸は、Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281 (1982)に記載されるものなどの様々な技術によって生体試料から抽出することができる。場合によっては、核酸は、最初に生体試料から抽出され、次いでインビトロでクロスリンクされ得る。場合によっては、天然の会合タンパク質（例えば、ヒストン）を、核酸からさらに除去することができる。

20

【0350】

他の実施形態では、本開示は、例えば、組織、細胞培養物、体液、動物組織、植物、細菌、真菌、ウイルスなどから単離されたDNAを含む任意の高分子量二本鎖DNAに容易に適用することができる。

【0351】

いくつかの実施形態では、複数の独立した試料の各々は、少なくとも約1 ng、2 ng、5 ng、10 ng、20 ng、30 ng、40 ng、50 ng、75 ng、100 ng、150 ng、200 ng、250 ng、300 ng、400 ng、500 ng、1 μg、1.5 μg、2 μg、5 μg、10 μg、20 μg、50 μg、100 μg、200 μg、500 μg、または1000 μg、またはそれ以上の核酸材料を独立して含むことができる。いくつかの実施形態では、複数の独立した試料の各々は、約1 ng、2 ng、5 ng、10 ng、20 ng、30 ng、40 ng、50 ng、75 ng、100 ng、150 ng、200 ng、250 ng、300 ng、400 ng、500 ng、1 μg、1.5 μg、2 μg、5 μg、10 μg、20 μg、50 μg、100 μg、200 μg、500 μg、または1000 μg未満、またはそれ以上の核酸を独立して含むことができる。

30

【0352】

いくつかの実施形態では、末端修復は、Epicentre Biotechnologies (Madison, WI) から入手可能なものなどの市販のキットを使用して、平滑末端5'リン酸化核酸末端を生成するために行われる。

40

【0353】

アダプター

アダプターオリゴヌクレオチドは、標的ポリヌクレオチドに結合することができる配列（少なくともその一部が公知である）を有する任意のオリゴヌクレオチドを含む。アダプターオリゴヌクレオチドは、DNA、RNA、ヌクレオチドアナログ、非カノニカルヌクレオチド、標識ヌクレオチド、修飾ヌクレオチド、またはそれらの組み合わせを含むことができる。アダプターオリゴヌクレオチドは、一本鎖、二本鎖、または部分二本鎖であり得る。一般に、部分二本鎖アダプターは、1つ以上の一本鎖領域および1つ以上の二本鎖

50

領域を含む。二本鎖アダプターは、互いにハイブリダイズした2つの別個のオリゴヌクレオチド（「オリゴヌクレオチド二本鎖」とも呼ばれる）を含むことができ、ハイブリダイゼーションは、1つ以上の平滑末端、1つ以上の3'オーバーハング、1つ以上の5'オーバーハング、ミスマッチおよび／または不対合のヌクレオチドから生じる1つ以上のバルジ、またはこれらの任意の組み合わせを残し得る。いくつかの実施形態では、一本鎖アダプターは、互いにハイブリダイズすることができる2つ以上の配列を含む。このような2つのハイブリダイズ可能な配列が一本鎖アダプターに含まれる場合、ハイブリダイゼーションによりヘアピン構造（ヘアピンアダプター）が得られる。アダプターの2つのハイブリダイズした領域がハイブリダイズされていない領域によって互いに隔てられている場合、「バブル」構造が生じる。バブル構造を含むアダプターは、内部ハイブリダイゼーションを含む単一のアダプターオリゴヌクレオチドからなり得るか、または互いにハイブリダイズされた2つ以上のアダプターオリゴヌクレオチドを含み得る。アダプターにおける2つのハイブリダイズ可能な配列の間などの内部配列ハイブリダイゼーションは、一本鎖アダプターオリゴヌクレオチドにおいて二本鎖構造を生じ得る。ヘアピンアダプターおよび二本鎖アダプターなどの異なる種類のアダプター、または異なる配列のアダプターを組み合わせることができる。ヘアピンアダプター中のハイブリダイズ可能な配列は、オリゴヌクレオチドの一方または両方の末端を含んでも含まなくてもよい。いずれの末端もハイブリダイズ可能な配列に含まれない場合、両方の末端は「フリー」または「オーバーハング」である。一方の末端のみがアダプター中の別の配列にハイブリダイズ可能である場合、他方の末端は、3'オーバーハングまたは5'オーバーハングなどのオーバーハングを形成する。5'末端ヌクレオチドおよび3'末端ヌクレオチドが相補的であり、互いにハイブリダイズするように、5'末端ヌクレオチドおよび3'末端ヌクレオチドの両方がハイブリダイズ可能な配列に含まれる場合、末端は「平滑」と呼ばれる。異なるアダプターを標的ポリヌクレオチドに連続反応でまたは同時に結合することができる。例えば、第1および第2のアダプターを同じ反応に加えることができる。アダプターは、標的ポリヌクレオチドと組み合わせる前に操作することができる。例えば、末端リン酸塩を添加または除去することができる。

10

20

30

40

50

【0354】

アダプターは、限定されないが、1つ以上の増幅プライマーアニーリング配列またはその相補体、1つ以上の配列決定プライマーアニーリング配列またはその相補体、1つ以上のバーコード配列、複数の異なるアダプターまたは異なるアダプターのサブセット間で共有される1つ以上の共通配列、1つ以上の制限酵素認識部位、1つ以上の標識ポリペプチドオーバーハングに相補的な1つ以上のオーバーハング、1つ以上のプローブ結合部位（例えば、Illumina, Inc.によって開発されたものなどの大規模並列配列決定のためのフローセルなどの配列決定プラットフォームへの結合のための）、1つ以上のランダムまたはほぼランダムな配列（例えば、1つ以上の位置において2つ以上の異なるヌクレオチドのセットからランダムに選択される1つ以上のヌクレオチドであって、1つ以上の位置において選択される異なるヌクレオチドのそれぞれが、ランダム配列を含むアダプターのプールにおいて表される、1つ以上のヌクレオチド）、およびそれらの組み合わせを含む。2つ以上の配列要素は、互いに隣接していなくてもよく（例えば、1つ以上のヌクレオチドによって分離される）、互いに隣接していてもよく、部分的に重複していても、完全に重複していてもよい。例えば、増幅プライマーアニーリング配列は、配列決定プライマーアニーリング配列としても役立つことができる。配列要素は、3'末端もしくはその付近に、5'末端もしくはその付近に、またはアダプターオリゴヌクレオチドの内部に位置することができる。アダプターオリゴヌクレオチドが、ヘアピンなどの二次構造を形成することができる場合、配列要素は、部分的にもしくは完全に二次構造の外側に、部分的にもしくは完全に二次構造の内側に、または二次構造に関与する配列の間に位置することができる。例えば、アダプターオリゴヌクレオチドがヘアピン構造を含む場合、配列要素は、ハイブリダイズ可能な配列（「ループ」）間の配列の中を含む、ハイブリダイズ可能な配列（「ステム」）の内側または外側に部分的または完全に位置することができ

る。いくつかの実施形態では、異なるバーコード配列を有する複数の第1のアダプターオリゴヌクレオチド中の第1のアダプターオリゴヌクレオチドは、複数の第1のアダプターオリゴヌクレオチドのすべての間で共通の配列要素を含む。いくつかの実施形態では、すべての第2のアダプターオリゴヌクレオチドは、第1のアダプターオリゴヌクレオチドによって共有される共通の配列要素とは異なる、すべての第2のアダプターオリゴヌクレオチド間で共通の配列要素を含む。配列要素の差異は、例えば、配列長の変化、1つ以上のヌクレオチドの欠失もしくは挿入、または1つ以上のヌクレオチド位置におけるヌクレオチド組成の変化（塩基変化または塩基修飾など）に起因して、異なるアダプターの少なくとも一部が完全にアラインメントしないような任意のものであり得る。いくつかの実施形態において、アダプターオリゴヌクレオチドは、1つ以上の標的ポリヌクレオチドに相補的な5'オーバーハング、3'オーバーハング、またはその両方を含む。相補的なオーバーハングは、限定されないが、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、またはそれ以上のヌクレオチド長を含む、1つ以上のヌクレオチド長であり得る。例えば、相補的なオーバーハングは、約1、2、3、4、5、または6ヌクレオチド長であり得る。相補的なオーバーハングは、固定配列を含み得る。相補的なオーバーハングは、1つ以上のヌクレオチドのランダム配列を含み得、1つ以上のヌクレオチドは、1つ以上の位置において2つ以上の異なるヌクレオチドのセットからランダムに選択され、1つ以上の位置において選択された異なるヌクレオチドの各々は、ランダム配列を含む相補的なオーバーハングを有するアダプターのプールにおいて表される。いくつかの実施形態において、アダプターオーバーハングはアデニンまたはチミンからなる。

10

20

【0355】

アダプターオリゴヌクレオチドは、それらが含まれる1つ以上の配列要素を収容するのに少なくとも十分な任意の適切な長さを有し得る。いくつかの実施形態では、アダプターは、約10、約15、約20、約25、約30、約35、約40、約45、約50、約55、約60、約65、約70、約75、約80、約90、約100、約200もしくはそれ以上、10未満、15未満、20未満、25未満、30未満、35未満、40未満、45未満、50未満、55未満、60未満、65未満、70未満、75未満、80未満、90未満、100未満、200もしくはそれ以上を下回る、または、10超、15超、20超、25超、30超、35超、40超、45超、50超、55超、60超、65超、70超、75超、80超、90超、100超、200もしくはそれ以上を上回るヌクレオチド長である。いくつかの例では、アダプターは、約10～約50ヌクレオチド長であり得る。さらなる例において、アダプターは、約20～約40ヌクレオチド長であり得る。

30

【0356】

本明細書で使用される場合、「バーコード」という用語は、バーコードが関連するポリヌクレオチドのいくつかの特徴を同定することを可能にする公知の核酸配列を指す。いくつかの実施形態では、同定されるポリヌクレオチドの特徴は、ポリヌクレオチドが由来する試料である。いくつかの実施形態では、バーコードは、少なくとも3、4、5、6、7、8、9、10、11、12、13、14、15、またはそれ以上のヌクレオチド長であり得る。例えば、バーコードは、少なくとも10、11、12、13、14、または15ヌクレオチド長であり得る。いくつかの実施形態において、バーコードは10、9、8、7、6、5、または4ヌクレオチド長より短くてもよい。例えば、バーコードは10ヌクレオチド長より短くてもよい。いくつかの実施形態では、いくつかのポリヌクレオチドに関連するバーコードは、他のポリヌクレオチドに関連するバーコードとは異なる長さである。一般に、バーコードは十分な長さであり、それらが関連するバーコードに基づく試料の同定を可能にするのに十分に異なる配列を含む。いくつかの実施形態では、バーコード、およびそれが関連する試料源は、バーコード配列中の1つ以上のヌクレオチドの突然変異、挿入、または欠失、例えば、1、2、3、4、5、6、7、8、9、10またはそれ以上のヌクレオチドの突然変異、挿入、または欠失などの後に、正確に同定することができる。いくつかの例では、1、2、または3個のヌクレオチドが突然変異、挿入、および/または欠失され得る。いくつかの実施形態では、複数のバーコード中の各バーコードは

40

50

、少なくとも2つのヌクレオチド位置、例えば、少なくとも2、3、4、5、6、7、8、9、10、またはそれ以上の位置において、複数のバーコード中の他のすべてのバーコードとは異なる。いくつかの例において、各バーコードは、少なくとも2、3、4、または5個の位置において、すべての他のバーコードとは異なり得る。いくつかの実施形態では、第1の部位および第2の部位の両方が、複数のバーコード配列のうちの少なくとも1つを含む。いくつかの実施形態では、第2の部位のバーコードは、第1のアダプターオリゴヌクレオチドのバーコードから独立して選択される。いくつかの実施形態では、バーコードを有する第1の部位および第2の部位は、対の配列が同じまたは異なる1つ以上のバーコードを含むように対にされる。いくつかの実施形態では、本開示の方法は、標的ポリヌクレオチドが結合されるバーコード配列に基づいて、標的ポリヌクレオチドが由来する試料を同定する工程をさらに含む。一般に、バーコードは、標的ポリヌクレオチドに結合されたときに、標的ポリヌクレオチドが由来する試料の識別子としての役割を果たす核酸配列を含み得る。

10

【0357】

アダプターオリゴヌクレオチドは、免疫グロブリンまたは免疫グロブリン結合タンパク質またはその断片に結合(coupled、linked)され得、またはテザリングされ得る。例えば、クロスリンクした試料をMNaseなどのDNaseで、インサイチュでゲノム消化した後、1つ以上の抗体を試料に添加して、メチル化部位または転写因子結合部位などで消化されたクロマチンに結合させることができる。次に、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合(coupled、linked)され、またはテザリングされたビオチン化アダプターオリゴヌクレオチドを試料に添加して、アダプターをクロマチン中の1つ以上の特異的部位に標的化することができる。次いで、試料をリガーゼで処理して、近接ライゲーションを行ってもよい。さらに、ストレプトアビジンを使用して、アダプターにライゲーションされたDNAを単離してもよい。次いで、クロスリンクは、PCRおよび配列決定を使用して試料を増幅する前に、脱クロスリンクされ得る。代替的に、アダプター連鎖オリゴヌクレオチドは、クリックケミストリーを使用して精製試薬に結合することができる修飾ヌクレオチドを含み得る。

20

【0358】

架橋オリゴヌクレオチド

本明細書で提供される方法は、接合部において複数のセグメントの第1のセグメントおよび第2のセグメントを結合させる工程を含むことができる。場合によっては、結合させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端をfish-inすること、および平滑末端をライゲーションすることを含むことができる。特定の場合において、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントを架橋オリゴヌクレオチドに接触させることを含むことができる。図15は、第1のセグメントおよび第2のセグメントを結合するために架橋オリゴヌクレオチドを使用する例示的なワークフローを示しており、第1のセグメントおよび第2のセグメントを形成するために核酸がインサイチュで消化される。架橋オリゴヌクレオチドを第1のセグメントおよび第2のセグメントの各々にライゲーションする前に、末端を平滑化し、ポリアデニル化する。次いで、第1のセグメントおよび第2のセグメントをライゲーションして、架橋オリゴヌクレオチドを含む接合部を作製する。様々な場合において、結合させる工程は、少なくとも第1のセグメントおよび第2のセグメントをバーコードに接触させることを含むことができる。

30

40

【0359】

いくつかの実施形態では、本明細書で提供される架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチド長～約50ヌクレオチド長であり得る。特定の実施形態では、架橋オリゴヌクレオチドは、少なくとも約15ヌクレオチド長～約18ヌクレオチド長であり得る。様々な実施形態において、架橋オリゴヌクレオチドは、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少な

50

くとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50またはそれ以上のヌクレオチド長であり得る。一例において、架橋オリゴヌクレオチドは、少なくとも10ヌクレオチド長である。別の例では、架橋オリゴヌクレオチドは、12ヌクレオチド長または約12ヌクレオチド長である。場合によっては、少なくとも10bpの架橋オリゴヌクレオチドは、安定性を増加させ、短い挿入物、染色体間ライゲーション、非特異的ライゲーション、および架橋自己ライゲーションなどの有害な近接ライゲーション事象を低減することができる。

【0360】

いくつかの実施形態において、架橋オリゴヌクレオチドは、バーコードを含み得る。特定の実施形態では、架橋オリゴヌクレオチドは、複数のバーコード（例えば、2つ以上のバーコード）を含むことができる。様々な実施形態において、架橋オリゴヌクレオチドは、共に結合された複数の架橋オリゴヌクレオチドを含むことができる。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断片に結合され（coupledまたはlinked）得る。場合によっては、結合された架橋オリゴヌクレオチドは、抗体が結合される試料核酸中の位置に送達され得る。

【0361】

分割およびプーリングアプローチを用いて、固有のバーコードを有する架橋オリゴヌクレオチドを生成することができる。試料の集団は、複数の群に分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドバーコードが群間で異なるが、群内では同じであるように試料に結合させることができ、試料の群を再度プールすることができ、このプロセスを複数回繰り返すことができる。例えば、ポリヌクレオチドの集団をA群とB群へ分割することができる。例えば、第1の架橋オリゴヌクレオチドは、A群のポリヌクレオチドに結合することができ、第2の架橋オリゴヌクレオチドは、B群のポリヌクレオチドに結合することができる。これにより、架橋オリゴヌクレオチドバーコードはA群内で同じではあるが、架橋オリゴヌクレオチドはA群とB群との間で異なる。このプロセスを反復すると、最終的には、集団中の各試料が固有の一連の架橋オリゴヌクレオチドバーコードを有することになり、単一試料（例えば、単細胞、単核、単染色体）の分析が可能になる。1つの例示的な例では、ビーズの固体支持体に結合したクロスリンクした消化核の試料を8本のチューブに分け、各々が、ライゲーションされる二本鎖DNA（dsDNA）アダプターを含む第1のアダプター群（第1の反復）の8つの固有のメンバーのうちの1つを含有する。8つのアダプターの各々は、核中のクロスリンクしたクロマチン凝集体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有することができるが、そうでなければ固有のdsDNA配列を有する。第1のアダプター群をライゲーションした後、核をプールして戻し、洗浄してライゲーション反応成分を除去することができる。分配、ライゲーション、およびプールのスキームは、さらに2回繰り返すことができる（2回の反復）。各アダプター群からのメンバーのライゲーション後に、クロスリンクしたクロマチン凝集体を複数のバーコードに連続して結合させることができる。場合によっては、複数のアダプター群の複数のメンバーの連続ライゲーション（反復）は、バーコードの組み合わせをもたらす。利用可能なバーコードの組み合わせの数は、反復当たりの群の数、および使用されるバーコードオリゴヌクレオチドの総数に依存する。例えば、それぞれ8つのメンバーを含む3つの反復は、83の可能な組み合わせを有することができる。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加または減少させることによって、および／または、反復の数を増加または減少させることによって調整することができる。1を超えるアダプター群が使用される場合、反復アダプター結合のために、分配、結合、およびプールのスキームを使用することができる。場合によっては、分配、結合、およびプールのスキームは、少なくとも3、4、5、6、7、8、9、または10回さらに繰り返され得る。場合によっては

10

20

30

40

50

、最後のアダプター群のメンバーは、例えば、PCR増幅による配列決定ライブラリー調製中に、アダプターに結合したDNAのその後の濃縮のための配列を含む。

【0362】

（分割およびプールの）このプロセスの反復は、最終的には、集団中の各試料が固有の一連の架橋オリゴヌクレオチドバーコードを有することになり、単一試料（例えば、単細胞、単核、および単染色体）の分析が可能になる。図16および図17は、核酸がインサイチュで消化され、その後、末端の平滑化およびポリアデニル化される、分割およびプールのアプローチを使用する例示的なワークフローを示す。単一細胞を分注し、バーコードを各細胞に存在する末端にライゲーションする（例えば、バーコードbc1）。細胞をプールし、その後、単細胞を単離し、第2のバーコードを各細胞に存在する末端にライゲーションする（例えば、バーコードbc2）。細胞を再度プールし、単一細胞に分離した後、架橋アダプター（例えば、Bio-Bridge）をライゲーションし、これは、バーコードおよびアダプターの固有の組み合わせを有する2つのセグメント間で接合部を形成する別のDNAセグメントにライゲーションすることができ、接合部が由来する細胞を同定する（例えば、バーコードbc1およびbc2）。架橋アダプターは、その後のプルダウンまたは他の精製のために、ビオチンなどの1つ以上の親和性試薬を含むことができる。図18は、分割およびプールのアプローチから生じるバーコードおよび架橋の組み合わせの例を示す。

10

【0363】

別の例示的な例では、ビーズの固体支持体に結合したクロスリンクした消化核の試料を8本のチューブに分けることができ、各々が、ライゲーションされる二本鎖DNA（dsDNA）アダプターを含む第1のアダプター群（第1の反復）の8つの固有のメンバーのうちの1つを含有する。8つのアダプターの各々は、核中のクロスリンクしたクロマチン凝集体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有することができるが、そうでなければ固有のdsDNA配列を有する。第1のアダプター群をライゲーションした後、核をプールして戻し、洗浄してライゲーション反応成分を除去することができる。分配、ライゲーション、およびプールのスキームは、さらに2回繰り返すことができる（2回の反復）。各アダプター群からのメンバーのライゲーション後に、クロスリンクしたクロマチン凝集体を複数のバーコードに連続して結合させることができる。

20

【0364】

場合によっては、複数のアダプター群の複数のメンバーの連続ライゲーション（反復）は、バーコードの組み合わせをもたらすことができる。利用可能なバーコードの組み合わせの数は、反復当たりの群の数、および使用されるバーコードオリゴヌクレオチドの総数に依存し得る。例えば、それぞれ8つのメンバーを含む3つの反復は、83の可能な組み合わせを有することができる。場合によっては、バーコードの組み合わせは固有である。特定の場合には、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加または減少させることによって、および／または、反復の数を増加または減少させることによって調整することができる。1を超えるアダプター群が使用される場合、反復アダプター結合のために、分配、結合、およびプールのスキームを使用することができる。様々な場合において、分配、結合、およびプールのスキームは、少なくとも3、4、5、6、7、8、9、10回またはそれ以上さらに繰り返され得る。場合によっては、最後のアダプター群のメンバーは、例えば、PCR増幅による配列決定ライブラリー調製中に、アダプターに結合したDNAのその後の濃縮のための配列を含み得る。

30

40

【0365】

場合によっては、分割-プール戦略を可能にする3つのオリゴ設計が使用され得、それによって8つの異なるビオチン化オリゴと組み合わせた2つの96ウェルプレートが使用され得、73, 728の異なる分子の別々のバーコード化が可能となる。特定の場合において、8つのオリゴの第1の2つのセットはビオチン化されず、8つのオリゴの第3のセットがビオチン化される。様々な場合において、各バーコード化オリゴヌクレオチドは方

50

向性があり、各ラウンドにおいて1つのオリゴのみを加えることを可能にする。架橋オリゴヌクレオチドは、対応する末端と適合することを可能にする配列を有することができる。

【0366】

特定の場合において、バーコードおよびアダプターは、完全にライゲーションされた架橋によって取られる配列空間の量を低減するために、より短い配列を有することもある。様々な場合において、架橋は、30bpの配列空間を占め得る。場合によっては、架橋は、54bpの配列空間を占め得るが、固有の分子識別子(UMI)のためのさらなる位置を提供し得る。特定の場合において、UMIは、73, 728の異なる組み合わせの単一細胞同定を可能にし得る。様々な場合において、第1の2つのオリゴセットは未修飾であり、第3のオリゴセットはビオチン化されている。

10

【0367】

架橋アダプター中のバーコード配列を使用して、試料の多重配列決定を可能にすることができる。例えば、近接ライゲーションは、いくつかの異なる試料に対して行うことができ、各試料は、異なるバーコード配列を有する架橋オリゴヌクレオチドを使用する。次いで、多重配列決定分析のために試料をプールすることができ、バーコード配列に基づいて個々の試料に配列情報を逆多重化することができる。

【0368】

ゲノムアセンブリおよびハプロタイプフェージングのためのフェージングされたリードセット

20

ロングリードまたはショートリード配列決定技術を使用して、ゲノムアセンブリおよびハプロタイプフェージングを含む用途のために、フェージングされたリードセットを含むリードセットを生成するための方法が本明細書で提供される。いくつかのそのような方法は、その全体が参照により本明細書に組み込まれるWO2017/147279においてより詳細に提供される。このような方法では、核酸分子を(例えば、クロマチン構造において)結合させ、切断して内部末端を露出させ、他の露出末端へ接合部で再結合させ、結合から解放し、配列決定することができる。この技術は、複数の配列セグメントを含む核酸分子を産生することができる。核酸分子内の複数の配列セグメントは、それらの天然の位置および配向または出発位置および配向に対して再編成されている間、保存されたフェージング情報を有することができる。接合部のいずれかの側の配列セグメントは、試料核酸分子の同じフェーズに由来すると確信してみなすことができる。

30

【0369】

高分子量DNAを含む核酸分子は、少なくとも1つの核酸結合部分上で結合または固定化され得る。例えば、インビトロのクロマチン凝集体へと組み立てられてホルムアルデヒド処理で固定されたDNAは、本明細書の方法に一貫している。核酸を結合または固定化するアプローチは、限定されないが、インビトロのまたは再構成クロマチンのアセンブリ、天然のクロマチン、DNA結合タンパク質凝集体、ナノ粒子、DNA結合物質を使用して被覆されたDNA結合ビーズ、ポリマー、合成DNA結合分子、または他の固体あるいはほぼ固体の親和性分子を含む。場合によっては、ビーズは、固相可逆的固定化(SPRI)ビーズ(例えば、Beckman-Coulter Agencourt AMPure XPビーズなどの負に帯電したカルボキシル基を含むビーズ)である。

40

【0370】

本明細書に記載されるものなどの核酸結合部分に結合される核酸は、配列決定デバイス上でのリード距離よりも長い距離(例えば、10kb、50kb、100kbまたはそれ以上)だけ核酸分子上で分離された第1のセグメントおよび第2のセグメントを持つ核酸分子が、それらの共通のリン酸ジエステル結合とは独立して共に結合されるように保持され得る。そのような結合された核酸分子の切断に際し、第1のセグメントおよび第2のセグメントの露出された末端は、互いに対しライゲーションすることができる。場合によっては、核酸分子は、固体表面上で結合された核酸分子間に重なりがほとんどまたはまったくないような濃度で結合され、そうすることで、切断された分子の露出した内部末端は再

50

びライゲーションするか、または、切断前に共通の核酸由来の相同性となる共通の核酸由来の相同性となる他のセグメントから露出した末端にのみ再び結合するようになる可能性が高い。結果的に、DNA分子は切断され、切断された露出した内部末端は、フェージング情報を失うことなく、例えば無作為に再びライゲーションすることができる。

【0371】

結合した核酸分子は、任意の数の酵素的および非酵素的アプローチのうちの1つによって内部末端を露出させるために、切断され得る。例えば、核酸分子は、一本鎖オーバーハングを残す制限エンドヌクレアーゼなどの制限酵素を用いて、消化することができる。例えば、MboI消化はこの目的に適しているが、他の制限エンドヌクレアーゼも企図される。制限エンドヌクレアーゼのリストは、例えば、ほとんどの分子生物学生成物カタログにおいて入手可能である。核酸切断のための他の非限定的な技術としては、トランスポザーゼ、タグメンテーション酵素複合体、トポイソメラーゼ、非特異的エンドヌクレアーゼ、DNA修復酵素、RNA誘導型ヌクレアーゼ、フラグメンターゼ、または代替酵素の使用が挙げられる。例えば、トランスポザーゼは、トランスポザーゼ送達オリゴヌクレオチド配列の結合によって特徴付けられる核酸中の配列非依存的切断を作り出すために、結合されていない左右の境界と組み合わせて使用することができる。機械的手段（例えば、超音波処理、せん断）、熱的手段（例えば、温度変化）、または電磁的手段（例えば、UV照射などの照射）を含む物理的手段も、切断をもたらすために使用することができる。

【0372】

この段階での核酸の固定化は、切断された核酸分子断片を物理的に近接した状態に保つことができ、そうすることで、最初の分子のフェージング情報が保存される。1つの核酸結合部分から得られる例示的なクロマチン凝集体を、図28に概略的に示す。例えば、クロマチン凝集体への固定化の利点は、共通の核酸分子の別個の領域が、それらのホスホジエステル骨格とは無関係に共に保持され得、その結果、それらのフェージング情報が、ホスホジエステル骨格の切断時に失われないということである。この利点はさらに、核酸分子が切断前に結合される代替的な足場によっても伝達される。

【0373】

任意選択で、再アニーリングおよび再ライゲーションを防ぐために、一本鎖「粘着性」末端オーバーハングを修飾する。例えば、粘着末端は、例えば、1つのヌクレオチドおよびポリメラーゼを添加することによって、部分的にfill-inされる（図29）。このようにして、一本鎖末端全体をfill-inすることはできないが、末端は、以前は相補的であった末端との再ライゲーションを防ぐように修飾される。5' GATC 5'-プライムオーバーハングを残すMboI消化の例では、グアノシンヌクレオチド三リン酸のみが添加される。これは、第1の相補的塩基（「C」）の「G」fill-inのみをもたらし、5' GATオーバーハングをもたらす。この工程は、遊離粘着末端を互いに再ライゲーションに不適合なものにするが、下流での用途のために粘着性末端を保存する。代替的に、平滑末端は、オーバーハングの完全なfill-in、平滑末端生成酵素による制限消化、一本鎖DNAエキソヌクレアーゼによる処理、または非特異的切断によって生成される。場合によっては、トランスポザーゼは、平滑末端または粘着末端を有するアダプター末端を、DNA分子の露出した内部末端に結合させるために使用される。

【0374】

任意選択で、「切断オリゴヌクレオチド」が導入される（図30）。この切断オリゴヌクレオチドは、切断／再ライゲーション部位をマークする。いくつかの切断オリゴヌクレオチドは、露出した核酸試料の内部末端上に生成された部分的にfill-inされたオーバーハングと適合する一本鎖オーバーハングを両方の末端に有する。切断オリゴヌクレオチドの例を以下に示す。場合によっては、一本鎖オーバーハングを有する二本鎖オリゴヌクレオチドは、ライゲーション中にコンカテマーを形成することができないように、その5'末端での5'リン酸除去などによって修飾される。代替的に、平滑切断オリゴヌクレオチドを使用するか、または別個の切断オリゴヌクレオチドを使用して切断部位をマークしない。いくつかのシステムにおいて、例えば、トランスポザーゼが使用される場合、切

断はトランスポソーム (transpososome) 境界配列の付加、その後の境界配列の互いへのライゲーションまたは切断オリゴへのライゲーションによって達成される。例示的な切断オリゴを以下に示す。しかしながら、代替的な切断オリゴは本明細書の開示と一致しており、配列、長さ、オーバーハングの存在もしくは配列、または 5' 脱リン酸化などの修飾が変化する。

【0375】

【化1】

5' ATCACGCGC 3'

3' TGC GCGCTA 5'

10

場合によっては、切断オリゴヌクレオチドの二本鎖領域は変化する。切断オリゴヌクレオチドの関連する特徴は、そのオーバーハングの配列であり、核酸試料へのライゲーションを可能にするが、任意選択で、自己ライゲーションまたはコンカテマー形成を妨げるように修飾される。しばしば、切断オリゴヌクレオチドは、下流の配列反応において容易に同定されるように、標的核酸分子中に存在しないかまたは存在しにくい配列を含むことが好ましい。切断オリゴは任意選択で、例えば、既知のバーコード配列またはランダムに生成された固有の識別子配列でバーコード化される。固有の識別子配列は、核酸分子中または試料中の複数の接合部が同じ固有の識別子でバーコード化される可能性を非常に低くするように設計することができる。

20

【0376】

切断された末端は、例えば、リガーゼまたは類似の酵素を用いて、直接またはオリゴ（例えば、切断オリゴ）を介して互いに結合することができる。固定化された高分子量の核酸分子の遊離一本鎖末端が直接または切断オリゴヌクレオチドにライゲーションされるように、ライゲーションを進行させることができる（図30）。切断オリゴヌクレオチドは、利用される場合、2つのライゲーション可能な末端を有することができるため、このライゲーションは、高分子量核酸分子の領域を効果的にまとめて鎖状にすることができる。2つの露出した末端間に切断配列または分子を付加するという結果をもたらす代替的なアプローチも、切断なしに2つの露出した末端を直接結合するためのアプローチと同様に用いることができる。

30

【0377】

次いで、核酸を核酸結合部分から遊離させることができる。インビトロクロマチン凝集体の場合、これは、クロスリンクを脱クロスリンクすること、またはタンパク質成分を消化すること、またはクロスリンクを脱クロスリンクすることとタンパク質成分を消化することの両方によって、達成することができる。適切なアプローチは、プロテイナーゼKによる複合体の処置であるが、多くの代替物も企図される。他の結合技術については、リンカー分子の切断または基質の分解などの適切な方法を用いることができる。

【0378】

40

このような技術から得られる核酸分子は、様々な関連する特徴を有することができる。核酸分子内の配列セグメントは、それらの天然または出発の位置および配向に対して再編成され得るが、フェージング情報は保存される。その結果、接合部のいずれかの側の配列セグメントは、共通の試料分子の共通フェーズに確実に割り当てることができる。したがって、分子上で互いに遠く離れたセグメントは、このような技術によって、各セグメントの部分または全体が単一分子配列決定デバイスの単一のランにおいて配列決定され、決定的なフェーズ割り当てを可能にするように、一緒にまたは近接させることができる。代替的に、場合によっては、最初に隣接するセグメントは、結果として生じる核酸において互いに分離され得る。場合によっては、核酸分子は、再ライゲーションのうち少なくとも約50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、

50

92%、93%、94%、95%、96%、97%、98%、99%、99.9%、99.99%、99.999%、または100%が、切断前に共通の核酸由来の相同性となるセグメント間になるように、再度ライゲーションされ得る。

【0379】

得られた分子の別の関連する特徴は、場合によっては、元の分子配列の大部分またはすべてが、おそらく再編成されているが、最終的な切断または再編成された分子において保存されることである。例えば、場合によっては、1%、2%、3%、4%、5%、10%、15%、または20%以下の元の分子が、結果として生じる分子（複数可）を生成する際に失われる。したがって、フェーズ決定因子として有用であることに加えて、結果として生じる分子は、元の分子配列の実質的な割合を保持し、その結果、結果として生じる分子は任意選択で、デノボ配列決定において有用なコンティグ情報などの配列情報を同時に生成するために、または以前に生成されたコンティグ情報の独立した検証として、使用される。

10

【0380】

いくつかの結果として生じる分子のライブラリーの別の特徴は、切断接合部が結果として生じる分子の集団の複数のメンバーには共通しないことである。すなわち、同じ出発核酸分子の異なるコピーは、接合部および再編成の異なるパターンで終わることができる。ランダム切断接合部は、非特異的切断分子を用いて、または制限エンドヌクレアーゼ選択もしくは消化パラメーターの変化を通して、生成することができる。

20

【0381】

分子特異的な切断部位を有することの結果は、場合によっては、切断オリゴヌクレオチドが任意選択で、悪影響を及ぼさないように「切断分子」の再シャッフリングおよび再ライゲーションをもたらすプロセスから除外されることである。3つ以上の再シャッフリングされた分子のセグメントをアラインメントさせることによって、切断部位がライブラリーの他のメンバーの大部分にそれらが存在しないことによって容易に同定されることが観察される。すなわち、3つ以上の再シャッフリングされた分子が局所的にアラインメントされている場合、セグメントはすべての分子に共通であることをわかるが、セグメントのエッジは分子間で異なり得る。セグメントの局所的配列類似性がどこで終わるかに注目することにより、「非切断 (unpunctuated)」再編成核酸分子中の切断接合部をマッピングすることができる。

30

【0382】

得られた核酸分子（例えば、図31を参照）は、例えば、ロングリードシーケンサーで配列決定することができる。得られた配列リードは、元の入力分子由来の核酸配列と、それらが使用される場合には、切断オリゴの配列との間で交互するセグメントを含有する。これらのリードは、切断オリゴヌクレオチド配列を使用して各リードからの配列データを分割するためにコンピュータによって処理されることができ、またはそうでなければ、接合部を同定するために処理される。各リード内の配列セグメントは、単一の入力高分子量DNA分子からのセグメントであり得る。元の核酸分子は、染色体などのゲノム配列またはその画分を含むことができる。セグメントリードのセットは、元の核酸分子において不連続であり得るが、ロングレンジのハプロタイプフェージングされたデータを明らかにする。これらのデータは、デノボゲノムアセンブリに、および入力ゲノム中のフェージングヘテロ接合位置のフェージングのために、使用することができる。接合部間の配列は、供給源核酸試料中の連続する核酸配列を示し、接合部を横切る配列は、核酸試料中で同相であるが、配列された足場では隣接セグメントから遠く離れて得る核酸セグメントを示している。

40

【0383】

接合部は、様々なアプローチによって同定することができる。切断オリゴが使用される場合、接合部は、切断オリゴ配列を含有するリードにおいて同定され得る。代替的に、接合部は、以前に生成されたコンティグ配列データセット、または独立して誘導された接合部を有する第2の独立して生成されたDNA鎖分子などの核酸分子の第2の配列供給源（

50

および、好ましくは、第3の配列供給源)との比較によって、同定することができる。配列がアラインメントされると、例えば、特定の位置へのアラインメントの質または信頼度は、1つのセグメントがどこで終了し、別のセグメントがどこで終わり、別のセグメントがどこで始まるかを示すことができる。切断をもたらすために制限酵素が使用される場合、制限酵素認識部位を含む配列は、潜在的に接合部を含むと評価され得る。いくつかの制限酵素認識部位は、酵素によって物理的にアクセス可能ではない場合があるため、すべての制限酵素認識部位が接合部を含むわけではなく、一方で、例えば、核酸は支持体に結合していたことに留意されたい。統計情報は、接合部を同定する際にも採用することができ、例えば、接合部間の長さセグメントは、特定の平均値であると予測されるか、または特定の分布に従うと予測され得る。

10

【0384】

本明細書における操作の利点は、分子の非隣接領域を近接させながら分子フェージング情報を保存することができ、そうすることで、分子の非隣接領域が、ロングリードなどの単一のリードにおける配列決定に適した距離で単一の核酸分子中に含まれることである。したがって、出発試料において単一のロングリード操作の距離(例えば、10 kb、15 kb、20 kb、30 kb、50 kb、100 kb、またはそれ以上)よりも大きく離れている領域は、それらがロングレンジ配列決定反応の単一のリードによってカバーされる距離内にあるように、局所的に近接する。したがって、元の試料中の単一のリードのための配列決定技術の距離を超えて分離される領域は、フェーズ保存され、再編成された分子中の単一の反応において読み取られる。

20

【0385】

得られた再構成された分子を配列決定することができ、それらの配列情報を、独立してもしくは同時に生成された配列リードもしくはコンティグ情報に、または既知の参照ゲノム配列(例えば、ヒトゲノムの既知の配列)にマッピングすることができる。得られた再編成された分子リード上で隣接するセグメントは相同性であると推定される。したがって、これらのセグメントが異種のコンティグまたはロングレンジ配列リードにマッピングされる場合、リードは、配列アセンブリにおける共通分子の共通フェーズに割り当てられる。

【0386】

代替的に、複数の独立して生成された、結果として生じる再編成された分子が同時に配列決定される場合、フェージングされた試料データは任意選択でこれらの分子のみから生成され、そうすることで接合部によって分離されるセグメント配列が相同性であると推測される一方で、接合部によって分離されない配列が、試料自体において連続する核酸のストレッチを表し、例えば、デノボ配列決定に有用であり、かつフェーズ決定に有用であると推測される。しかしながら、付加的または代替的に、同時に配列決定された複数の独立して生成された結果として生じる再編成された分子は、独立して生成された足場またはコンティグ情報と依然として比較され得る。

30

【0387】

本明細書に提示される方法および組成物は、特に、配列決定技術におけるリードの長さ(例えば、10 kb、20 kb、50 kb、100 kb、500 kb、またはそれ以上)よりも長い距離だけ分離された分子セグメントに関して、ロングレンジフェージング情報を保存することができるが、セグメントが単一のリードによってカバーされるのに十分なほど隣接しているかまたは近接している再編成されたまたはしばしば「切断された」分子におけるそのような非隣接セグメントを提供する。

40

【0388】

いくつかの例では、結果として生じる再編成された分子は、配列決定のために天然の分子と組み合わせられる。天然の分子は、使用される場合、切断配列の欠如によって情報的に認識および利用することができる。天然の分子は、短いまたはロングリード技術を使用して配列決定され、それらのアセンブリは、再編成された分子またはライブラリーの配列決定を通して生成されるフェージング情報およびセグメント配列情報によって誘導される。

50

【0389】

切断オリゴヌクレオチド

場合によっては、切断オリゴヌクレオチドは、露出した切断末端を結合する際に使用され得る。切断オリゴヌクレオチドは、フェーズを保存する再編成をしている試料分子の2つの切断された内部末端をクロスリンクするように、標的ポリヌクレオチドに結合され得る任意のオリゴヌクレオチドを含んでいる。切断オリゴヌクレオチドは、DNA、RNA、ヌクレオチドアナログ、非カノニカルヌクレオチド、標識されたヌクレオチド、修飾されたヌクレオチド、またはそれらの組み合わせを含み得る。多くの例において、二本鎖切断オリゴヌクレオチドは、互いにハイブリダイズされた2つの別個のオリゴヌクレオチド（「オリゴヌクレオチド二重鎖」とも称される）を含み、ハイブリダイゼーションは、1つ以上の平滑末端、1つ以上の3'オーバーハング、1つ以上の5'オーバーハング、ミスマッチおよび／または非対合のヌクレオチドから結果として生じる1つ以上のバルジ、またはこれらの任意の組み合わせを残し得る。いくつかの例において、異なる切断オリゴヌクレオチドは、連続的な反応においてまたは同時に、ポリヌクレオチドを標的とするために結合される。例えば、第1および第2の切断オリゴヌクレオチドが同じ反応に加えられ得る。代替的に、切断オリゴの集団は場合によっては均一である。

10

【0390】

切断オリゴヌクレオチドを、標的ポリヌクレオチドと組み合わせる前に操作することができる。例えば、末端リン酸塩が除去され得る。そのような修飾は、試料分子の切断された内部末端ではなく、互いに対する切断オリゴの位置を排除する。

20

【0391】

切断オリゴヌクレオチドは、様々な配列要素のうち1つ以上を含んでおり、限定されないが、配列またはその補体をアニーリングする1つ以上の増幅プライマー、配列またはその補体をアニーリングする1つ以上の配列決定プライマー、1つ以上のバーコード配列、多くの異なる切断オリゴヌクレオチドまたは異なる切断オリゴヌクレオチドのサブセットで共有される1つ以上の共通配列、1つ以上の制限酵素認識部位、1つ以上の標的ポリヌクレオチドオーバーハングに相補的な1つ以上のオーバーハング、1つ以上のプローブ結合部位、1つ以上のランダムまたはほぼランダムな配列、およびそれらの組み合わせを含む。いくつかの例において、2つ以上の配列要素は、互いに隣接しておらず（例えば、1つ以上のヌクレオチドにより分離される）、または、部分的もしくは完全に重なって互いに隣接している。例えば、配列をアニーリングする増幅プライマーは、配列をアニーリングする配列決定プライマーとしても役立つ。特定の例において、配列要素は、3'末端またはその付近に、5'末端またはその付近に、あるいは切断オリゴヌクレオチドの内部に位置付けられる。

30

【0392】

代替的な実施形態において、切断オリゴは、配列決定反応において占められる配列情報の量を最小化するように二本鎖分子の完全性を維持するための塩基の最小の補体を含むか、または、切断オリゴはライゲーションのための最適な数の塩基を含むか、または切断オリゴの長さは恣意的に決定される。

【0393】

いくつかの実施形態において、切断オリゴヌクレオチドは、1つ以上の標的ポリヌクレオチドに相補的である、5'オーバーハング、3'オーバーハング、またはその両方を含む。特定の例において、相補的なオーバーハングは、1以上のヌクレオチド長であり、限定されないが、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、またはそれ以上のヌクレオチド長を含む。例えば、相補的なオーバーハングは、約1、2、3、4、5、または6ヌクレオチド長である。いくつかの実施形態において、切断オリゴヌクレオチドのオーバーハングは、制限エンドヌクレアーゼ消化または他のDNA切断方法によって生成された標的ポリヌクレオチドのオーバーハングに相補的である。

40

【0394】

切断オリゴヌクレオチドは、それらが構成される1つ以上の配列要素を収容するのに少

50

なくとも十分な任意の適切な長さを有することができる。いくつかの実施形態において、切断オリゴヌクレオチドは、約 4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100、200、もしくはそれ以上のヌクレオチド長、約 4 未満、5 未満、6 未満、7 未満、8 未満、9 未満、10 未満、15 未満、20 未満、25 未満、30 未満、35 未満、40 未満、45 未満、50 未満、55 未満、60 未満、65 未満、70 未満、75 未満、80 未満、90 未満、100 未満、200 未満、もしくはそれ以上のヌクレオチド長、または、約 4 超、5 超、6 超、7 超、8 超、9 超、10 超、15 超、20 超、25 超、30 超、35 超、40 超、45 超、50 超、55 超、60 超、65 超、70 超、75 超、80 超、90 超、100 超、200 超、もしくはそれ以上のヌクレオチド長である。いくつかの例において、切断オリゴヌクレオチドは、5～15ヌクレオチド長である。さらなる例において、切断オリゴヌクレオチドは約 20～約 40ヌクレオチドである。

10

【0395】

好ましくは、切断オリゴヌクレオチドは、例えば、5'リン酸の切除（アルカリンホスファターゼ処理を介して、またはそのような部分がない状態で合成により新規に）によって修飾され、その結果、多量体を形成するために互いにライゲーションしない。3'OH（ヒドロキシル）部分は、切断された核酸上で5'リン酸にライゲーションすることができ、それにより第1または第2の核酸セグメントに対するライゲーションを支持する。

【0396】

アダプターオリゴヌクレオチド

20

アダプターは、標的ポリヌクレオチドに結合され得る配列を有する任意のオリゴヌクレオチドを含む。様々な例において、アダプターオリゴヌクレオチドは、DNA、RNA、ヌクレオチドアナログ、非カノニカルヌクレオチド、標識されたヌクレオチド、修飾されたヌクレオチド、またはそれらの組み合わせを含む。いくつかの例において、アダプターオリゴヌクレオチドは、一本鎖、二本鎖、または部分的に二重である。一般に、部分的に二重のアダプターオリゴヌクレオチドは、1つ以上の一本鎖領域および1つ以上の二本鎖領域を含む。二本鎖アダプターオリゴヌクレオチドは、互いにハイブリダイズされた2つの別個のオリゴヌクレオチド（「オリゴヌクレオチド二重鎖」とも称される）を含み、ハイブリダイゼーションは、1つ以上の平滑末端、1つ以上の3'オーバーハング、1つ以上の5'オーバーハング、ミスマッチおよび／または非対合のヌクレオチドから結果として生じる1つ以上のバルジ、またはこれらの任意の組み合わせを残し得る。いくつかの実施形態において、一本鎖アダプターオリゴヌクレオチドは、互いにハイブリダイズすることができる2つ以上の配列を含む。2つのそのようなハイブリダイズ可能な配列が一本鎖アダプターに含まれていると、ハイブリダイゼーションはヘアピン構造（ヘアピンアダプター）を産生する。アダプターオリゴヌクレオチドの2つのハイブリダイズされた領域がハイブリダイズされていない領域によって互いに分離されると、「バブル」構造が結果として生じる。バブル構造を含むアダプターオリゴヌクレオチドは、内部のハイブリダイゼーションを含む単一のアダプターオリゴヌクレオチドからなるか、または互いにハイブリダイズされた2つ以上のアダプターオリゴヌクレオチドを含む。アダプターオリゴヌクレオチドにおける2つのハイブリダイズ可能な配列間などでの内部配列ハイブリダイゼーションは、いくつかの例において、一本鎖アダプターオリゴヌクレオチド中に二本鎖構造を生成する。いくつかの例において、異なる種類のアダプターオリゴヌクレオチドは、異なる配列のヘアピンアダプターおよび二本鎖アダプターなどの組み合わせで使用される。特定の場合では、ヘアピンアダプター中のハイブリダイズ可能な配列は、オリゴヌクレオチドの末端の一方または両方を含んでいる。いずれの末端もハイブリダイズ可能な配列に含まれていない場合、両方の末端は「遊離（free）」または「オーバーハング」である。一方の末端のみがアダプター中で別の配列にハイブリダイズ可能な場合、他の末端は3'オーバーハングまたは5'オーバーハングなどのオーバーハングを形成する。5'-末端ヌクレオチドおよび3'-末端ヌクレオチドの両方がハイブリダイズ可能な配列に含まれ、それにより5'-末端ヌクレオチドおよび3'-末端ヌクレオチドが相補的にかつ互い

30

40

50

にハイブリダイズ可能となる場合、末端は「平滑」と称される。場合によっては、異なるアダプターオリゴヌクレオチドは、連続的な反応においてまたは同時に、ポリヌクレオチドを標的とするために結合される。例えば、第1および第2のアダプターオリゴヌクレオチドは同じ反応に加えられる。いくつかの例において、アダプターオリゴヌクレオチドは標的ポリヌクレオチドと組み合わせる前に操作される。例えば、末端リン酸が付加または除去され得る。

【0397】

アダプターオリゴヌクレオチドは、様々な配列要素のうちの1つ以上を含み、限定されないが、配列またはその補体をアニーリングする1つ以上の増幅プライマー、配列またはその補体をアニーリングする1つ以上の配列決定プライマー、1つ以上のバーコード配列、多数の異なるアダプターまたは異なるアダプターのサブセットで共有される1つ以上の共通配列、1つ以上の制限酵素認識部位、1つ以上の標的ポリヌクレオチドオーバーハングに相補的な1つ以上のオーバーハング、1つ以上のプローブ結合部位（例えば、Illumina, Inc.により開発されたものなど、大規模並列配列決定のためのフローセルなどの配列決定プラットフォームへの結合のためのもの）、1つ以上のランダムまたはほぼランダムな配列（例えば、1つ以上の位置で2つ以上の異なるヌクレオチドのセットからランダムに選択された1つ以上のヌクレオチドであり、異なるヌクレオチドの各々はランダム配列を含むアダプターのプールの中で表わされる1つ以上の位置で選択される）、およびこれらの組み合わせを含む。多くの例において、2つ以上の配列要素は、互いに隣接しておらず（例えば、1つ以上のヌクレオチドにより分離される）、または、部分的もしくは完全に重なって互いに隣接し得る。例えば、配列をアニーリングする増幅プライマーは、配列をアニーリングする配列決定プライマーとしても役立つ。配列要素は、3'末端もしくはその付近に、5'末端もしくはその付近に、またはアダプターオリゴヌクレオチドの内部に位置する。アダプターオリゴヌクレオチドがヘアピンなどの二次構造を形成することができる場合、配列要素は、二次構造の外部に部分的もしくは完全に、二次構造の内部に部分的もしくは完全に、または二次構造に関与する配列間に位置することができる。例えば、アダプターオリゴヌクレオチドがヘアピン構造を含む場合、配列要素は、ハイブリダイズ可能な配列（「ループ」）間の配列中を含む、ハイブリダイズ可能な配列（「ステム」）の内部または外部に部分的または完全に位置することができる。いくつかの実施形態において、異なるバーコード配列を有する複数の第1のアダプターオリゴヌクレオチドにおける第1のアダプターオリゴヌクレオチドは、複数におけるすべての第1のアダプターオリゴヌクレオチド中で共通の配列要素を含む。いくつかの実施形態において、すべての第2のアダプターオリゴヌクレオチドは、第1のアダプターオリゴヌクレオチドにより共有される共通配列要素とは異なる、すべての第2のアダプターオリゴヌクレオチドに共通の配列要素を含む。配列要素の差異は、異なるアダプターの少なくとも一部が、例えば、配列長の変化、1つ以上のヌクレオチドの欠失もしくは挿入、または1つ以上のヌクレオチド位置でのヌクレオチド組成の変化（塩基の変化または塩基の修飾など）が原因で完全にはアラインメントしないようなものであり得る。いくつかの実施形態において、アダプターオリゴヌクレオチドは、1つ以上の標的ポリヌクレオチドに相補的である、5'オーバーハング、3'オーバーハング、またはその両方を含む。相補的なオーバーハングは、1以上のヌクレオチド長であり、限定されないが、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、またはそれ以上のヌクレオチド長を含む。例えば、相補的なオーバーハングは、約1、2、3、4、5、または6ヌクレオチド長であり得る。相補的なオーバーハングは固定化された配列を含み得る。相補的なオーバーハングは、1つ以上のヌクレオチドのランダム配列を付加的または代替的に含んでもよく、その結果、1つ以上のヌクレオチドが1つ以上の位置で2つ以上の異なるヌクレオチドのセットからランダムに選択され、異なるヌクレオチドの各々は、ランダム配列を含む相補的なオーバーハングを有するアダプターオリゴヌクレオチドのプールで表わされる1つ以上の位置で選択される。いくつかの実施形態において、アダプターオリゴヌクレオチドのオーバーハングは、制限エンドヌクレアーゼ消化によって生成された標的ポリヌクレ

10

20

30

40

50

オチドのオーバーハングに相補的である。いくつかの実施形態において、アダプターオリゴヌクレオチドのオーバーハングはアデニンまたはチミンからなる。

【0398】

アダプターオリゴヌクレオチドは、それらが構成される1つ以上の配列要素を収容するのに少なくとも十分な任意の適切な長さを有することができる。いくつかの実施形態において、アダプターオリゴヌクレオチドは、約4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100、200、もしくはそれ以上のヌクレオチド長、約4未満、5未満、6未満、7未満、8未満、9未満、10未満、15未満、20未満、25未満、30未満、35未満、40未満、45未満、50未満、55未満、60未満、65未満、70未満、75未満、80未満、90未満、100未満、200未満、もしくはそれ以上のヌクレオチド長、または、約4超、5超、6超、7超、8超、9超、10超、15超、20超、25超、30超、35超、40超、45超、50超、55超、60超、65超、70超、75超、80超、90超、100超、200超、もしくはそれ以上のヌクレオチド長である。いくつかの例において、アダプターオリゴヌクレオチドは、5～15ヌクレオチド長である。さらなる例において、アダプターオリゴヌクレオチドは約20～約40ヌクレオチド長である。

10

【0399】

好ましくは、アダプターオリゴヌクレオチドは、例えば、5'リン酸の切除（アルカリンホスファターゼ処理を介して、またはそのような部分がない状態で合成により新規に）によって修飾され、その結果、多量体を形成するために互いにライゲーションしない。3'OH（ヒドロキシル）部分は、切断された核酸上で5'リン酸にライゲーションすることができ、それにより第1または第2の核酸セグメントに対するライゲーションを支持する。

20

【0400】

核酸試料のフェージング情報の判定

核酸試料のフェージング情報を判定するために、核酸は、例えば、本明細書で議論される抽出方法によって最初に獲得される。多くの場合、核酸はその後、核酸分子の切断の後にフェージング情報を維持するように固体表面に結合される。好ましくは、核酸分子は、再構成クロマチンを生成するために核酸結合タンパク質と共にインビトロで組み立てられるが、他の適切な固体表面は、核酸結合タンパク質凝集体、ナノ粒子、核酸結合ビーズ、または、核酸結合物質、ポリマー、合成核酸結合分子、あるいは他の固体またはほぼ固体の親和性分子で被覆されたビーズを含む。核酸試料は、天然のクロマチンの場合など固体表面にすでに結合された状態で得ることができる。天然のクロマチンは、ホルマリン固定／パラフィン包埋（FFPE）されたまたは同様に維持された試料の形態などですでに固定されている状態で得ることができる。

30

【0401】

核酸結合部分への結合の後、結合された核酸分子は切断され得る。切断は、任意の数の酵素方法および非酵素方法を含む、任意の適切な核酸切断実体で実行される。好ましくは、DNA切断は、制限エンドヌクレアーゼ、フラグメンターゼ（fragmentase）、またはトランスポザーゼで実行される。代替的または付加的に、核酸切断は、他の制限酵素、トポイソメラーゼ、非特異的エンドヌクレアーゼ、核酸修復酵素、RNA誘導型ヌクレアーゼ、または代替的な酵素で達成される。機械的手段（例えば、超音波処理、せん断）、熱的手段（例えば、温度変化）、または電磁的手段（例えば、UV照射などの照射）を含む物理的手段も、切断をもたらすために使用することができる。核酸切断は、使用される切断方法に依存して、「粘着性」オーバーハングまたは平滑末端のいずれかを有する遊離核酸末端を産生する。粘着性オーバーハング末端が生成されると、粘着末端は任意選択で、再ライゲーションを妨げるために部分的にfill-inされる。代替的に、オーバーハングは、平滑末端を産生するために完全にfill-inされる。

40

【0402】

50

多くの場合、オーバーハング末端は、d N T Pで部分的にまたは完全に f i l l e d i n され、これは任意選択で標識される。そのような場合、d N T Pは、ビオチン化され、硫酸化され、フルオロフォアに結合され、脱リン酸化され、または他の任意の数のヌクレオチド修飾を受けることができる。ヌクレオチド修飾は、メチル化（例えば、5 - m C、5 - h m C、5 - f C、5 - c a C、4 - m C、6 - m A、8 - o x o G、8 - o x o A）などのエピジェネティック修飾を含み得る。標識または修飾は、ナノ細孔配列決定によって検出可能なエピジェネティック修飾などの、配列決定中に検出可能なものから選択することができる。このように、ライゲーション結合の位置は配列決定中に検出することができる。これらの標識または修飾は、結合または濃縮のために標的とされ得、例えば、メチル-シトシンを標的とする抗体は、メチル-シトシンで f i l l e d i n された平滑末端を捕捉し、標的化し、結合し、または標識するために使用され得る。非天然ヌクレオチド、非標準ヌクレオチド、または修飾されたヌクレオチド、および核酸アナログも、平滑末端の f i l l e d i n の位置を標識するために使用され得る。非カノニカルヌクレオチドまたは修飾されたヌクレオチドは、プソイドウリジン（Ψ）、ジヒドロウリジン（D）、イノシン（I）、7-メチルグアノシン（m7G）、キサンチン、ヒポキサンチン、プリン、2,6-ジアミノプリン、および6,8-ジアミノプリンを含み得る。核酸アナログは、ペプチド核酸（PNA）、モルホリノおよびロックド核酸（LNA）、グリコール核酸（GNA）、およびトレオース核酸（TNA）を含み得る。場合によっては、オーバーハングは、ビオチンのないd N T Pなどの、標識されていないd N T Pで f i l l e d i n される。場合によっては、トランスポゾンでの切断など、f i l l e d i n を必要としない平滑末端が生成される。トランスポザーゼが2つの結合されていない切断オリゴヌクレオチドを挿入場合、このような遊離平滑末端が生成される。しかしながら、切断オリゴヌクレオチドは、要望通りに粘着末端または平滑末端を有するように合成される。ヒストンなどの試料核酸に関連付けられるタンパク質も修飾され得る。例えば、ヒストンは、アセチル化（例えば、リジン残基にて）および／またはメチル化（例えば、リジンとアルギニン残基にて）され得る。

10

20

30

40

50

【0403】

次に、切断された核酸分子が依然として固体表面に結合されている一方、遊離核酸末端は共に連鎖される。連鎖は、場合によっては、ライゲーションを通じて、遊離末端間で、またはオリゴヌクレオチドなどの別個の実体で生じる。場合によっては、オリゴヌクレオチドは切断オリゴヌクレオチドである。そのような場合、切断分子末端は、切断された核酸分子の遊離末端と適合性を有する。多くの場合、切断分子はオリゴヌクレオチドのコンカテマー化（concatemerization）を防ぐために脱リン酸化される。大半の場合、切断分子は、切断された核酸分子の遊離核酸末端へと各末端でライゲーションされる。多くの場合、このライゲーション工程の結果、切断された核酸分子の再編成がもたらされ、その結果、出発核酸分子において互いに本来隣接していなかった2つの遊離末端はここでペアエンドで結合される。

【0404】

切断された核酸分子の遊離末端の結合の後、再編成された核酸試料は、任意の数の標準的な酵素方法および非酵素方法を使用して、核酸結合実体から放たれる。例えば、インビトロの再構成クロマチンの場合、再編成された核酸分子は、核酸結合タンパク質を変性または分解することによって放たれる。他の例において、クロスリンクは脱クロスリンクされる。また他の例において、親和性相互作用は脱クロスリンクまたは遮断される。放たれた核酸分子は、入力核酸分子と比較して再編成される。切断分子が使用される場合、結果として生じる再編成された分子は、再編成された核酸分子の全体にわたって散在する切断オリゴヌクレオチドにより、切断された分子（punctuated molecule）と称される。これらの場合、切断部（punctuations）に隣接する核酸セグメントはペアエンドを構築する。

【0405】

本明細書に開示される方法の切断および結合の工程中に、核酸分子がこれらのプロセス

の全体にわたって固体表面に結合されるため、フェージング情報は維持される。これにより、一塩基変異多型（SNP）などの他のマーカーからの情報に依存することなく、フェージング情報の分析が可能とされ得る。本明細書に開示される方法および組成物を使用して、場合によっては、核酸分子内の2つの核酸セグメントは、元々の核酸分子上に存在していたよりも近くにあるように、再編成される。多くの例において、出発核酸試料における2つの核酸セグメントの元々の分離距離は、標準配列決定技術の平均的なリード長よりも大きい。例えば、入力核酸試料内の2つの核酸セグメント間の出発分離距離は、約10 kb、12.5 kb、15 kb、17.5 kb、20 kb、25 kb、30 kb、35 kb、40 kb、45 kb、50 kb、60 kb、70 kb、80 kb、90 kb、100 kb、125 kb、150 kb、200 kb、300 kb、400 kb、500 kb、600 kb、700 kb、800 kb、900 kb、1 Mb、またはそれ以上である。好ましい例において、再編成された2つのDNAセグメント間の分離距離は、標準配列決定技術の平均的なリード長未満である。例えば、再編成されたDNA分子内の2つの再編成されたDNAセグメントを分離する距離は、約50 kb、40 kb、30 kb、25 kb、20 kb、17 kb、15 kb、14 kb、13 kb、12 kb、11 kb、10 kb、9 kb、8 kb、7 kb、6 kb、5 kb未満、またはそれ以下である。好ましい場合、分離距離は、ロングリード配列決定機械の平均的なリード長未満である。これらの場合において、再編成されたDNA試料が核酸結合部分から放たれ、配列決定されると、フェージング情報が判定され、デノボの配列足場を生成するのに十分な配列情報が生成される。

10

20

30

40

50

【0406】

再編成された核酸分子のバーコード化

いくつかの例において、本明細書に記載される放たれた再編成核酸分子は、配列決定前にさらに処理される。例えば、再編成された核酸分子内に含まれる核酸セグメントは、バーコード化され得る。バーコード化は、配列リードのより容易なグループ化を可能にし得る。例えば、バーコードは、同じ再編成された核酸分子から生じる配列を同定するために使用され得る。バーコードは個々の接合部を固有に同定するためにも使用され得る。例えば、各接合部は、接合部を固有に同定することができる固有の（例えば、ランダムに生成された）バーコードでマークを付けることができる。同じ再編成された核酸分子から生じる配列を同定するための第1のバーコード、および個々の接合部を固有に同定するための第2のバーコードなど、複数のバーコードを共に使用することができる。

【0407】

バーコード化は多くの技術を介して実現され得る。場合によっては、バーコードは切断オリゴ内の配列として含まれ得る。他の場合、放たれた再編成核酸分子は、少なくとも2つのセグメントを含むオリゴヌクレオチドに接触させることができ、1つのセグメントはバーコードを含み、別のセグメントは切断配列に相補的な配列を含む。切断配列へとアニーリングした後、バーコード化されたオリゴヌクレオチドは、ポリメラーゼで拡張されて、同じ切断された核酸分子からバーコード化された分子を生成する。切断された核酸分子は、フェージング情報が保存されている入力核酸分子の再編成されたバージョンであるため、生成されたバーコード化された分子も同じ入力核酸分子から由来する。これらのバーコード化された分子は、バーコード配列、切断相補的配列、およびゲノム配列を含む。

【0408】

切断によりまたは切断なしで再編成された核酸分子について、分子は他の手段によってバーコード化され得る。例えば、再編成された核酸分子は、再編成された核酸分子からの配列を組み込むために拡張され得る、バーコード化されたオリゴヌクレオチドに接触され得る。バーコードは、切断配列に、制限酵素認識部位に、関心対象の部位（例えば、ゲノム関心領域）に、またはランダム部位（例えば、バーコードオリゴヌクレオチド上の無作為のn量体配列を介して）に、ハイブリダイズすることができる。再編成された核酸分子は、試料中の他の再編成された核酸分子から適切な濃縮および／または分離（例えば、空間的または時間的な分離）を使用してバーコードに接触させることができ、その結果、複数の再編成された核酸分子は与えられず、同じバーコード配列が与えられる。例えば、再

編成された核酸分子を含む溶液は、1つの再編成された核酸分子のみが、所定のバーコード配列を伴ってバーコードまたはバーコードの群に接触されるような濃度に希釈され得る。バーコードは、遊離溶液中で、流体パーティション中で（例えば、液滴もしくはウェル）、またはアレイ上で（例えば、特定のアレイスポットにて）、再編成された核酸分子に接触され得る。

【0409】

バーコード化された核酸分子（例えば、伸長生成物）は、例えば、ショートリード配列決定機械上で配列決定することができ、フェージング情報は、共通のフェーズへの同じバーコードを有する配列のリードのグループ化によって判定される。代替的に、配列決定の前に、バーコード化された生成物は、例えば、ロングリード配列決定技術を使用して配列決定される長い分子を生成するために、例えば、バルクライゲーション（bulk ligation）を介して共に結合され得る。これらの場合、埋め込まれたリード対は、増幅アダプターおよび切断配列を介して同定可能である。さらに詳しいフェージング情報はリード対のバーコード配列から得られる。

【0410】

ペアエンドでのフェージング情報の判定

本明細書には、ペアエンドからフェージング情報を判定するための方法および組成物が提供される。ペアエンドは、開示された方法、または提供された実施例にさらに例示されるもののいずれかによって生成され得る。例えば、後に切断される固体表面に結合される核酸分子の場合、遊離末端の再ライゲーションの後、再びライゲーションされた核酸セグメントは、例えば、制限消化によって固相が結合した核酸分子から放たれる。この放出の結果、複数のペアエンドがもたらされる。場合によっては、ペアエンドは増幅アダプターにライゲーションされ、増幅され、および短距離技術で配列決定される。これらの場合、複数の異なる核酸結合部分に結合した核酸分子からのペアエンドは、配列決定された試料内にある。しかしながら、ペアエンドの接合部のいずれかの側について、接合部に隣接する配列は共通の分子の共通のフェーズに由来することが、明確に結論付けられる。ペアエンドが切断オリゴヌクレオチドに結合される場合、配列決定リードにおけるペアエンド接合部は、切断オリゴヌクレオチド配列によって同定される。他の場合、ペアエンドは修飾されたヌクレオチドによって結合され、これは、使用される修飾されたヌクレオチドの配列に基づいて同定され得る。

【0411】

代替的に、ペアエンドの放出後、フリーのペアエンドは、増幅アダプターにライゲーションされ、増幅され得る。これらの場合、複数のペアエンドはその後、ロングリード配列決定技術を使用して読み取られる長い分子を生成するために、共にバルクライゲーションされる。他の例において、放出されたペアエンドは、増幅工程を介入させることなく互いにバルクライゲーションされる。いずれの場合も、包埋されたリード対は、切断配列または修飾されたヌクレオチドなどの、連鎖配列に隣接している天然のDNA配列を介して同定可能である。連鎖状のペアエンドは、長い配列デバイス上で読み取られ、複数の接合部のための配列情報が得られる。ペアエンドは複数の異なる核酸結合部分に結合したDNA分子から由来するため、増幅アダプター配列に隣接するものなどの2つの個々のペアエンドに及ぶ配列は、複数の異なるDNA分子にマッピングされることがわかる。しかしながら、ペアエンド接合部のいずれかの側について、接合部に隣接する配列は共通の分子の共通のフェーズに由来することが、明確に結論付けられる。例えば、切断された分子に由来するペアエンドの場合、切断配列に隣接する配列は、共通のDNA分子へと明確に割り当てられる。好ましい場合、本明細書に開示される方法および組成物を使用して個々のペアエンドが連鎖されるため、単一リードにおいて複数のペアエンドを配列決定することができる。

【0412】

特定の図を参照すると、本明細書で論じるいくつかの実施形態の図が見られる。図28において、切断された再編成されたフェーズ保存核酸分子を構築するプロセスにおける中

間体が見られる。単一の核酸分子は、再構成されたクロマチン複合体などの核酸結合部分に結合され、複合体をクロスリンクするためにホルムアルデヒドと接触させられている。複合体は、核酸分子の内部ループのみがクラスターから突出するように、集合的に再構成クロマチンと呼ばれる核酸結合成分と共にクラスターを形成する単一の核酸出発分子を含む。図28に示すように、突出ループは、粘着末端を生成するために、制限エンドヌクレアーゼMboIを用いて切断される。

【0413】

代替的な実施形態では、核酸分子は、SPRIでコーティングされたビーズまたは他の核酸結合剤でコーティングされたビーズなどのビーズまたは表面に結合される。核酸試料は、ビーズあたり1つの核酸分子のみが結合するような条件下で、または結合した核酸が切断後にクロスライゲーションする可能性が低いような条件下で結合する。さらに、切断は、別の制限エンドヌクレアーゼ、トランスポザーゼ、タグメンテーション酵素、非特異的エンドヌクレアーゼ、トポイソメラーゼ、またはエンドヌクレアーゼ活性を有する他の薬剤を用いて交互に達成される。

【0414】

図29では、図28の切断された核酸複合体が、オーバーハングの単一の位置をfill-inするように、核酸ポリメラーゼおよびdGTPの単一の集団を使用して処理されることが分かる。fill-in工程は、複合体の粘着末端が後の工程でクロスアニーリングおよびライゲーションするのを防止する。場合によっては、工程は除外され、複合体は、切断オリゴなしでクロスライゲーションされる。代替的に、平滑末端が生成されるか、または制限エンドヌクレアーゼではなくトランスポザーゼの作用を介してタグメンテーションアダプターが付加される。

【0415】

図30は、複合体の露出末端への切断オリゴのアニーリングおよびライゲーション後の図28および図29の複合体を示す。切断オリゴは、核酸塩基配列としてではなく、薄い実線として描かれている。切断オリゴは、例えば、5'リン酸基の除去によって、コンカテマー化を排除するように任意選択で修飾される。切断オリゴは任意選択で、図29で修飾されるように、遊離粘着末端と適合性であるように設計される。他の実施形態では、切断された核酸末端は、切断オリゴを介在させることなく、互いに直接ライゲーションすることができる。

【0416】

図31は、プロテイナーゼKを用いた処理による再構成されたクロマチンからのクロスリンクおよび遊離の脱クロスリンク後の放出された切断された核酸分子を描いている。最終産物の切断された核酸は、切断オリゴ401によって分離されたセグメント400を含む。セグメントは、元の核酸分子のフェージング情報を保存するが、ランダムに順序付けられ、開始分子に対して配向される。元の核酸分子の配列の実質的にすべてが、切断された分子中に存在するため、切断された分子の配列決定は、デノボコンティグを生成するのに十分な配列情報を生成する。

【0417】

ロングリード配列決定デバイスを使用して、切断された核酸を配列決定すると、局所的な順序および配向、ならびにフェージング情報が導出される、切断されていないセグメントに対応する配列のストレッチが観察される。切断オリゴ配列にわたる長い配列リードの領域も観察される。切断オリゴのいずれかの側のこれらの配列セグメントは、互いに相同性である（および、切断された分子上の他のセグメントと相同性である）ことが知られているが、正しい順序および配向にある可能性は低い。再編成プロセスの利点は、試料分子上の互いに遠く離れたセグメントが、単一の読み取りにおいて広がるように近接させられることである。別の利点は、デノボコンティグ情報が同時に生成されるように、元の試料分子の配列情報が大部分保存されることである。

【0418】

核酸

10

20

30

40

50

真核生物において、ゲノムDNAは、核内の染色体として構成されるクロマチンに詰められる。クロマチンの基礎的な構造単位は、ヒストン八量体の周囲に包まれるDNAの146塩基対(bp)からなるヌクレオソームである。ヒストン八量体は、コアヒストンH2A-H2B二量体およびH3-H4二量体の各々の2つのコピーからなる。ヌクレオソームは、共通して「数珠玉構造(beads on a string)」と称されるものにおいてDNAに沿って規則的に間隔を空けられる。

【0419】

ヌクレオソームへのコアヒストンおよびDNAのアセンブリは、シャペロンタンパク質および関連するアセンブリ因子により媒介される。これら因子のほぼすべては、コアヒストン結合タンパク質である。ヌクレオソームアセンブリタンパク質-1(NAP-1)などのヒストンシャペロンの一部は、ヒストンH3およびH4に結合する嗜好性を示す。新しく合成されたヒストンがアセチル化され、その後、クロマチンへのアセンブリ後に脱アセチル化されることも観察された。したがって、ヒストンのアセチル化または脱アセチル化を媒介する因子は、クロマチンアセンブリプロセスにおいて重要な役割を果たす。

【0420】

一般に、2つのインビトロ方法が、クロマチンを再構成またはアセンブルするために開発されている。1つの方法はATPに依存しないものであるが、もう一方はATPに依存するものである。クロマチンを再構成するためのATP非依存性の方法は、ヒストンシャペロンとして作用するようにタンパク質様NAP-1または塩のいずれかと、DNAまたはコアヒストンを含んでいる。この方法の結果、細胞中の天然のコアヌクレオソーム粒子を正確に模倣しないDNA上のヒストンのランダムな配置がもたらされる。これらの粒子は、規則的に順序づけられ、拡張されたヌクレオソームアレイではなく、使用されるDNA配列が通常は250bpよりも長くないため、しばしばモノヌクレオソームと呼ばれる(Kundu, T. K. et al., Mol. Cell 6: 551-561, 2000)。より大きな長さのDNA配列上の順序づけられたヌクレオソームの拡張アレイを生成するために、ATP依存性のプロセスを通じてクロマチンをアセンブルすることができる。

【0421】

過ヨウ素酸ヌクレオソームアレイのATP依存性アセンブリは、天然のクロマチンに見られるものと同様であり、DNA配列、コアヒストン粒子、シャペロンタンパク質、およびATPを利用するクロマチンアセンブリ因子を必要とする。ACF(ATPを利用するクロマチンアセンブリ因子およびリモデリング因子)またはRSF(リモデリングおよびスペーシングの因子)は、インビトロでクロマチンへのヌクレオソームの拡張され、順序づけられたアレイを生成するために使用される、2つの広く研究されたアセンブリ因子である(Fyodorov, D. V., and Kadonaga, J. T. Method Enzymol. 371: 499-515, 2003; Kundu, T. K. et al. Mol. Cell 6: 551-561, 2000)。

【0422】

特定の実施形態において、本開示の方法は、限定されないが、例えば、血漿、血清、および／または尿から単離された遊離DNA、細胞および／または組織からのアポトーシスDNA、(例えば、DNase Iによって)インビトロで酵素的に断片化されたDNAを含む、断片化された二本鎖DNAのあらゆる型に容易に適用することができる。

【0423】

生体試料から得られた核酸は、分析に適切な断片を生成するために断片化することができる。様々な酵素法を使用して、鋳型核酸を所望の長さに断片化してもよい。DNAはDNaseへの簡易な曝露によってランダムにせん断され得る。RNAは、RNase、熱とマグネシウムへの簡単な曝露、またはせん断によって、断片化され得る。RNAはcDNAへ変換され得る。フラグメンテーションが利用される場合、RNAはフラグメンテーションの前または後に、cDNAへ変換されてもよい。核酸分子は、一本鎖、二本鎖、または、一本鎖領域を含む二本鎖(例えば、ステム構造とループ構造)でもよい。

【0424】

いくつかの実施形態では、クロスリンクしたDNA分子は、サイズ選択工程に供され得る。核酸のサイズ選択は、特定のサイズ未満または特定のサイズを超えるクロスリンクしたDNA分子に対して行われ得る。サイズ選択は、クロスリンクの頻度および／またはフラグメンテーション方法によって、さらに影響され得る。いくつかの実施形態において、組成物は、約145bp～約600bp、約100bp～約2500bp、約600～約2500bp、約350bp～約1000bpの範囲、またはこれらの値のいずれかに囲まれる任意の範囲（例えば、約100bp～約2500bp）のDNA分子をクロスリンクすることを含んで、調製され得る。

【0425】

いくつかの実施形態において、試料のポリヌクレオチドは、1を超える特定の大きさの範囲の断片化DNA分子の集団へと断片化される。いくつかの実施形態において、断片は、少なくとも約1、約2、約5、約10、約20、約50、約100、約200、約500、約1000、約2000、約5000、約10,000、約20,000、約50,000、約100,000、約200,000、約500,000、約1,000,000、またはそれ以上のゲノム当量の出発DNAから生成され得る。フラグメンテーションはDNase処理によって達成され得る。いくつかの実施形態において、断片は、平均の長さが約10～約10,000、約20,000、約30,000、約40,000、約50,000、約60,000、約70,000、約80,000、約90,000、約100,000、約150,000、約200,000、約300,000、約400,000、約500,000、約600,000、約700,000、約800,000、約900,000、約1,000,000、約2,000,000、約5,000,000、約10,000,000、またはそれ以上のヌクレオチドである。いくつかの実施形態では、断片は、平均の長さが約145bp～約600bp、約100bp～約2500bp、約600～約2500bp、約350bp～約1000bp、またはこれらの値のいずれかに囲まれる任意の範囲（例えば、約100bp～約2500bp）である。いくつかの実施形態では、断片は、平均の長さが約2500bp未満、約1200bp未満、約1000bp未満、約800bp未満、約600bp未満、約350bp未満、または約200bp未満である。他の実施形態では、断片は、平均の長さが約100bp超、約350bp超、約600bp超、約800bp超、約1000bp超、約1200bp超、または約2000bp超である。DNaseの非限定的な例としては、DNase I、DNase II、小球菌ヌクレアーゼ、それらのバリエーション、およびそれらの組み合わせが挙げられる。例えば、DNase Iによる消化は、Mg⁺⁺がない状態およびMn⁺⁺がある状態で、DNA中のランダムな二本鎖切断を誘導することができる。フラグメンテーションは、5'オーバーハング、3'オーバーハング、平滑末端、またはそれらの組み合わせを有する断片を生成することができる。いくつかの実施形態において、前記方法は、アガロースゲルからのカラム精製または単離などの標準的な方法を介して、断片のサイズを選択する工程を含む。

【0426】

標的化されたヌクレアーゼ酵素

本明細書で提供される断片化DNAは、任意の数のヌクレアーゼ（例えば、制限エンドヌクレアーゼ）またはDNases（例えば、MNase）によるインサイチュ消化などの消化によって作製または生成することができる。場合によっては、酵素を組み合わせることで、所望の消化またはフラグメンテーションを達成することができる。様々な場合において、ヌクレアーゼ（またはそのドメインもしくは断片）は、1つ以上の抗体を使用して、特定のゲノム部位に標的化され得る。例えば、クロスリンクした試料は、ヒストン結合部位、転写因子結合部位、またはメチル化DNA部位などのDNAの特定の領域に結合する抗体と接触させることができる。その後、プロテインA、プロテインG、プロテインA/G、またはプロテインLなどの免疫グロブリン結合タンパク質またはその断

10

20

30

40

50

片に結合または融合したヌクレアーゼを試料に加えることができ、ヌクレアーゼは、抗体が結合した領域においてのみDNAを消化することができる。これは、例えば、所望の消化パターンを達成するために、第1の抗体がDNA試料に結合され、次いでヌクレアーゼが第1の抗体に標的化され、次いで第2の抗体がDNA試料に結合され、ヌクレアーゼが第2の抗体に標的化されるなど、組み合わせで行われ得る。

【0427】

ライゲーション

いくつかの実施形態において、断片化DNAの5'末端および／または3'末端のエンドヌクレオチド配列は、ライゲーション前に修飾されない。例えば、予測可能な平滑末端を残す酵素による切断を行い、その後、平滑末端化されたDNA断片の、核酸、例えば、平滑末端を含むアダプター、オリゴヌクレオチド、またはポリヌクレオチドへのライゲーションが行われ得る。いくつかの実施形態において、断片化DNA分子は、アダプターに結合される前に平滑末端を有するDNA断片を生成するために、平滑末端の平滑化（*blunt-end polished*）（または「末端修復される」）。平滑末端の平滑化の工程は、例えば、T4ポリメラーゼについて、3'～5'のエキソヌクレアーゼ活性および5'～3'のポリメラーゼ活性を有するDNAポリメラーゼなどの適切な酵素でのインキュベーションにより達成され得る。いくつかの実施形態において、末端修復の後に、オーバーハングを生成するために、1以上のアデニン、1以上のチミン、1以上のグアニン、または1以上のシトシンなどの1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、またはそれ以上のヌクレオチドの付加が続いて行われ得る。例えば、末端修復の後に、1、2、3、4、5、または6個のヌクレオチドの付加が続いて行われ得る。オーバーハングを有するDNA断片は、ライゲーション反応などにおいて、相補的なオーバーハングを有するオリゴヌクレオチド、アダプターオリゴヌクレオチド、またはポリヌクレオチドなどの1つ以上の核酸に結合され得る。例えば、単一のアデニンは、鑄型の独立したポリメラーゼを使用して末端修復されたDNA断片の3'末端に付加され、その後、3'末端で各々がチミンを有する1つ以上のアダプターへのライゲーションが行われ得る。いくつかの実施形態において、オリゴヌクレオチドまたはポリヌクレオチドなどの核酸は、1以上のヌクレオチドでの3'末端の伸長、その後の5'リン酸化により修飾された平滑末端の二本鎖DNA分子に結合され得る。場合によっては、3'末端の伸長は、マグネシウムを含有することができる適切な緩衝液中の1以上のdNTPの存在下で、Klenowポリメラーゼもしくは本明細書で提供される適切なポリメラーゼのいずれかなどのポリメラーゼにより、または、末端のデオキシヌクレオチドトランスフェラーゼの使用により、実行され得る。いくつかの実施形態において、平滑末端を有する標的ポリヌクレオチドは、平滑末端を含む1以上のアダプターに結合される。DNA断片分子の5'末端のリン酸化は、例えば、ATPおよびマグネシウムを含有する適切な緩衝液においてT4ポリヌクレオチドキナーゼにより実行され得る。断片化DNA分子は任意選択で、例えば、ホスファターゼなどの酵素の使用により、5'末端または3'末端を脱リン酸化するために処理され得る。

【0428】

本明細書で使用されるような「結合する（*connecting*および*joining*）」、および「ライゲーション」は、アダプターオリゴヌクレオチドおよび標的ポリヌクレオチドなどの2つのポリヌクレオチドに関して、隣接する骨格を有する単一のより大きなポリヌクレオチドを生成するための、2つの別個のDNAセグメントの共有結合を指す。2つのDNAセグメントを結合する方法は、限定されないが、酵素的および非酵素的（例えば、化学的）な方法を含む。非酵素的であるライゲーション反応の例は、引用により本明細書に組み込まれる、米国特許第5,780,613号と第5,476,930号に記載される非酵素的ライゲーション技術を含む。いくつかの実施形態において、アダプターオリゴヌクレオチドは、リガーゼ、例えば、DNAリガーゼまたはRNAリガーゼによって標的ポリヌクレオチドに結合される。各々が特徴化された反応条件を有する複数のリガーゼは、限定されないが、tRNAリガーゼ、Taq DNAリガーゼ、好熱菌DNA

リガーゼ、大腸菌 DNA リガーゼ、Tth DNA リガーゼ、サームス・スコトダクタス (Thermus scotoductus) DNA リガーゼ (I および II)、熱安定リガーゼ、Ampligase 熱安定 DNA リガーゼ、Vanc 型リガーゼ、9°N DNA リガーゼ、Tsp DNA リガーゼ、およびバイオプロスペクティングにより発見された新しいリガーゼを含む、NAD+ 依存性リガーゼ、T4 RNA リガーゼ、T4 DNA リガーゼ、T3 DNA リガーゼ、T7 DNA リガーゼ、Pfu DNA リガーゼ、DNA リガーゼ 1、DNA リガーゼ III、DNA リガーゼ IV、およびバイオプロスペクティングにより発見された新しいリガーゼを含む、ATP 依存性リガーゼ、ならびに、それらの野生型、突然変異体アイソフォーム、および遺伝的に操作されたバリエーションを含む。

10

【0429】

ライゲーションは、相補的なオーバーハングなどのハイブリダイズ可能な配列を有する DNA セグメント間で行われ得る。ライゲーションは 2 つの平滑末端間でも行われ得る。通常、5'リン酸塩がライゲーション反応に利用される。5'リン酸塩は標的ポリヌクレオチド、アダプターオリゴヌクレオチドまたは両方によって提供され得る。5'リン酸塩は、必要に応じて、結合される DNA セグメントに付加され得、または上記 DNA セグメントから除去され得る。5'リン酸塩の付加または除去のための方法は、限定されないが、酵素学的および化学的なプロセスを含む。5'リン酸塩の付加および／または除去に役立つ酵素は、キナーゼ、ホスファターゼ、およびポリメラーゼを含む。いくつかの実施形態において、ライゲーション反応（例えば、アダプター末端および標的ポリヌクレオチド末端）における結合された 2 つの末端の両方とも、2 つの末端を結合する際に 2 つの共有結合が作られるように、5'リン酸塩を提供する。いくつかの実施形態において、ライゲーション反応（例えば、アダプター末端および標的ポリヌクレオチド末端）における結合された 2 つの末端の 1 つのみが、2 つの末端を結合する際に 1 つの共有結合のみが作られるように、5'リン酸塩を提供する。

20

【0430】

いくつかの実施形態において、標的ポリヌクレオチドの一本鎖または両鎖での一鎖のみが、アダプターオリゴヌクレオチドに結合される。いくつかの実施形態において、標的ポリヌクレオチドの 1 つの末端または両方の末端における両鎖は、アダプターオリゴヌクレオチドに結合される。いくつかの実施形態において、3'リン酸塩はライゲーションの前に除去される。いくつかの実施形態では、アダプターオリゴヌクレオチドは、標的ポリヌクレオチドの両方の末端に付加され、各末端の 1 つまたは両方の鎖が、1 つ以上のアダプターオリゴヌクレオチドに結合されている。両方の末端の両方の鎖がアダプターオリゴヌクレオチドに結合されている場合、結合の後に、対応する 3'末端の伸長のための鋳型としての役割を果たすことができる 5'末端オーバーハングを残す切断反応が起き、その 3'末端はアダプターオリゴヌクレオチドに由来する 1 つ以上のヌクレオチドを含むこともある。いくつかの実施形態において、標的ポリヌクレオチドは、1 つの末端で第 1 のアダプターオリゴヌクレオチドに、もう 1 つの末端で第 2 のアダプターオリゴヌクレオチドに結合される。いくつかの実施形態において、標的ポリヌクレオチドの 2 つの末端は、単一のアダプターオリゴヌクレオチドの対向末端に結合される。いくつかの実施形態において、標的ポリヌクレオチドおよびそれが結合されるアダプターオリゴヌクレオチドは、平滑末端を含む。いくつかの実施形態において、別個のライゲーション反応は、少なくとも 1 つのバーコード配列を含む異なる第 1 のアダプターオリゴヌクレオチドを使用して各試料について実行され得、その結果、いかなるバーコード配列も 1 を超える試料の標的ポリヌクレオチドには結合されない。DNA セグメント、またはそれに結合したアダプターオリゴヌクレオチドを有する標的ポリヌクレオチドは、結合されたアダプターにより「タグ付けされた」と考慮される。

30

40

【0431】

場合によっては、ライゲーション反応は、約 0.1 ng/μL、約 0.2 ng/μL、約 0.3 ng/μL、約 0.4 ng/μL、約 0.5 ng/μL、約 0.6 ng/μL、

50

約 0.7 ng/μL、約 0.8 ng/μL、約 0.9 ng/μL、約 1.0 ng/μL、約 1.2 ng/μL、約 1.4 ng/μL、約 1.6 ng/μL、約 1.8 ng/μL、約 2.0 ng/μL、約 2.5 ng/μL、約 3.0 ng/μL、約 3.5 ng/μL、約 4.0 ng/μL、約 4.5 ng/μL、約 5.0 ng/μL、約 6.0 ng/μL、約 7.0 ng/μL、約 8.0 ng/μL、約 9.0 ng/μL、約 10 ng/μL、約 15 ng/μL、約 20 ng/μL、約 30 ng/μL、約 40 ng/μL、約 50 ng/μL、約 60 ng/μL、約 70 ng/μL、約 80 ng/μL、約 90 ng/μL、約 100 ng/μL、約 150 ng/μL、約 200 ng/μL、約 300 ng/μL、約 400 ng/μL、約 500 ng/μL、約 600 ng/μL、約 800 ng/μL、または約 1000 ng/μL の DNA セグメントまたは標的のポリヌクレオチドの濃度で実行され得る。例えば、ライゲーションは、約 100 ng/μL、約 150 ng/μL、約 200 ng/μL、約 300 ng/μL、約 400 ng/μL、または約 500 ng/μL の DNA セグメントまたは標的ポリヌクレオチドの濃度で実行され得る。

10

【0432】

場合によっては、ライゲーション反応は、約 0.1~1000 ng/μL、約 1~1000 ng/μL、約 1~800 ng/μL、約 10~800 ng/μL、約 10~600 ng/μL、約 100~600 ng/μL、または約 100~500 ng/μL の DNA セグメントまたは標的ポリヌクレオチドの濃度で実行され得る。

【0433】

場合によっては、ライゲーション反応は、約 5 分、約 10 分、約 20 分、約 30 分、約 40 分、約 50 分、約 60 分、約 90 分、約 2 時間、約 3 時間、約 4 時間、約 5 時間、約 6 時間、約 8 時間、約 10 時間、約 12 時間、約 18 時間、約 24 時間、約 36 時間、約 48 時間、または約 96 時間よりも長い間、実行され得る。他の場合、ライゲーション反応は、約 5 分未満、約 10 分未満、約 20 分未満、約 30 分未満、40 分未満、約 50 分未満、約 60 分未満、約 90 分未満、約 2 時間未満、約 3 時間未満、約 4 時間未満、約 5 時間未満、約 6 時間未満、約 8 時間未満、約 10 時間未満、約 12 時間未満、約 18 時間未満、約 24 時間未満、約 36 時間未満、約 48 時間未満、または約 96 時間未満にわたり実行され得る。例えば、ライゲーション反応は約 30 分~約 90 分の間、実行され得る。いくつかの実施形態において、標的ポリヌクレオチドに対するアダプターの結合は、アダプター由来のヌクレオチド配列を含む 3' オーバーハングを有する結合された産物のポリヌクレオチドを産生する。

20

30

【0434】

いくつかの実施形態において、標的ポリヌクレオチドに少なくとも 1 つのアダプターオリゴヌクレオチドを結合した後、1 つ以上の標的ポリヌクレオチドの 3' 末端は、鑄型として 1 つ以上の結合されたアダプターオリゴヌクレオチドを使用して伸長される。例えば、標的ポリヌクレオチドの 5' 末端のみに結合される 2 つのハイブリダイズされたオリゴヌクレオチドを含むアダプターは、鑄型としてアダプターの結合された鎖を使用して、標的の結合されていない 3' 末端の伸長を可能にし、それと同時にまたはその後、結合されていない鎖の変位を可能にする。2 つのハイブリダイズされたオリゴヌクレオチドを含むアダプターの両方の鎖は、標的ポリヌクレオチドに結合され得、結合された産物は 5' オーバーハングを有し、相補的な 3' 末端は鑄型として 5' オーバーハングを使用することで伸長され得る。さらなる例として、ヘアピンアダプターオリゴヌクレオチドは、標的ポリヌクレオチドの 5' 末端に結合され得る。いくつかの実施形態において、伸長される標的ポリヌクレオチドの 3' 末端は、アダプターオリゴヌクレオチドからの 1 つ以上のヌクレオチドを含む。アダプターが両方の末端で結合される標的ポリヌクレオチドについて、伸長は、5' オーバーハングを有する二本鎖標的ポリヌクレオチドの 3' 末端の両方に対し行われ得る。この 3' 末端の伸長、または「fill-in」反応は、鑄型へハイブリダイズされるアダプターオリゴヌクレオチドの鑄型に、相補的な配列または「補体」を生成し、それにより、二本鎖配列領域を生成するために 5' オーバーハングを fill-in する。二本鎖標的ポリヌクレオチドの両方の末端が相補的な鎖の 3' 末端の伸長により f

40

50

i l l - i nされる5'オーバーハングを有する場合、産物は完全に二本鎖である。伸長は、DNAポリメラーゼなどの任意の適切なポリメラーゼにより行うことができ、その多くが市販で入手可能である。DNAポリメラーゼは、DNA依存性DNAポリメラーゼ活性、RNA依存性DNAポリメラーゼ活性、またはDNA依存性およびRNA依存性のDNAポリメラーゼ活性を含み得る。DNAポリメラーゼは熱安定性または非熱安定性であり得る。DNAポリメラーゼの例は、限定されないが、Taqポリメラーゼ、Tthポリメラーゼ、Tliポリメラーゼ、Pfuポリメラーゼ、Pfu t u b oポリメラーゼ、Pyr o b e s tポリメラーゼ、Pwoポリメラーゼ、KODポリメラーゼ、Bstポリメラーゼ、Sacポリメラーゼ、Ssoポリメラーゼ、Pocポリメラーゼ、Pabポリメラーゼ、Mthポリメラーゼ、Phoポリメラーゼ、ES4ポリメラーゼ、VENTポリメラーゼ、DEEPVENTポリメラーゼ、EX-Taqポリメラーゼ、LA-Taqポリメラーゼ、Expandポリメラーゼ、Platinum Taqポリメラーゼ、Hi-Fiポリメラーゼ、Tbrポリメラーゼ、Tflポリメラーゼ、Truポリメラーゼ、Tacポリメラーゼ、Tneポリメラーゼ、Tmaポリメラーゼ、Tihポリメラーゼ、Tfiポリメラーゼ、Klenow断片、ならびにそれらのバリエーション、修飾した産物、および誘導体を含み、3'末端の伸長は、独立した試料から標的ポリヌクレオチドをプールする前または後に実行され得る。

10

【0435】

標的濃縮

特定の実施形態において、本開示は、標的核酸の濃縮および標的核酸の解析のための方法を提供する。場合によっては、濃縮の方法は溶液ベースのフォーマットである。場合によっては、標的核酸は標識剤で標識され得る。他の場合では、標的核酸は標識剤で標識される1つ以上の会合分子にクロスリンクされ得る。標識剤の例は、限定されないが、ビオチン、ポリヒスチジンタグ、および化学タグ（例えば、クリックケミストリー方法に使用されるアルキンおよびアジドの誘導体）を含む。さらに、標識された標的核酸は捕捉することができ、それにより捕捉剤の使用によって濃縮することができる。捕捉剤は、ストレプトアビジンおよび／またはアビジン、抗体、化学部分（例えば、アルキン、アジド）、および親和性精製に使用される任意の生物学的、化学的、物理的、または酵素的な薬剤であり得る。

20

【0436】

場合によっては、固定化されたまたは固定化されていない核酸プローブが、標的核酸を捕捉するために使用され得る。例えば、標的核酸は、固体支持体上または溶液中でのプローブへのハイブリダイゼーションにより、試料から濃縮され得る。いくつかの例において、試料はゲノム試料であり得る。いくつかの例において、プローブはアンプリコンであり得る。アンプリコンはあらかじめ定められた配列を含み得る。さらに、ハイブリダイズされた標的核酸は、プローブから洗浄および／または溶出され得る。標的核酸はDNA、RNA、cDNA、またはmRNAの分子であり得る。

30

【0437】

場合によっては、濃縮方法は、標的核酸を含む試料をプローブに接触させる工程、および標的核酸を固体支持体に結合させる工程を含み得る。場合によっては、試料は、標的核酸を得るための酵素的な方法を使用して、断片化され得る。場合によっては、プローブは、標的核酸へと特異的にハイブリダイズされ得る。場合によっては、標的核酸の平均サイズは、約145bp～約600bp、約100bp～約2500bp、約600～約2500bp、または約350bp～約1000bpであり得る。標的核酸はさらに、試料中の結合されていない核酸から分離され得る。固体支持体は、濃縮された標的核酸を得るために洗浄および／または溶出され得る。いくつかの例において、濃縮工程は約1、2、3、4、5、6、7、8、9、または10回繰り返され得る。例えば、濃縮工程は約1、2、または3回繰り返され得る。

40

【0438】

場合によっては、濃縮方法は、プローブ由来のアンプリコンを提供する工程を含むこと

50

ができ、増幅のための前記プローブは固体支持体に結合されている。固体支持体は、試料から特定の標的核酸を捕捉するために支持体に固定された核酸プローブを含み得る。プローブ由来のアンプリコンは、標的核酸にハイブリダイズされ得る。プローブアンプリコンへのハイブリダイゼーションの後、試料中の標的核酸は、捕捉されたプローブからハイブリダイズされた標的核酸を捕捉（例えば、ビオチン、抗体などの捕捉剤を介して）し、ならびに洗浄および／または溶出することにより、濃縮され得る。標的核酸配列はさらに、例えば、濃縮されたPCR産物の増幅されたプールを産生するためのPCR方法を使用して、増幅され得る。

【0439】

場合によっては、固体支持体は、マイクロアレイ、スライド、チップ、マイクロウェル、カラム、チューブ、粒子、またはビーズであり得る。いくつかの例において、固体支持体は、ストレプトアビジンおよび／またはアビジンで被覆され得る。他の例において、固体支持体は抗体で被覆され得る。さらに、固体支持体は、ガラス、金属、セラミック、またはポリマー材料を含み得る。いくつかの実施形態において、固体支持体は核酸マイクロアレイ（例えば、DNAマイクロアレイ）であり得る。他の実施形態において、固形支持体は常磁性ビーズであり得る。

【0440】

特定の実施形態において、本開示は、濃縮されたDNAを増幅する方法を提供する。場合によっては、濃縮されたDNAはリード対である。リード対は、本開示の方法により得ることができる。

【0441】

いくつかの実施形態において、1つ以上の増幅および／または複製の工程が、配列決定されるライブラリーの調製のために使用される。任意の適切な増幅方法が使用され得る。使用され得る増幅技術の例は、限定されないが、定量PCR、定量蛍光PCR（QF-PCR）、多重蛍光PCR（MF-PCR）、リアルタイムPCR（RT-PCR）、単細胞PCR、制限断片長多型PCR（PCR-RFLP）、PCK-RFLP-IRTP-PCR-IRFLP、ホットスタートPCR、ネステッドPCR、インサイチュポロニーPCR、インサイチュローリングサークル増幅（RCA）、架橋PCR、ライゲーションで媒介されたPCR、Qbレプリカーゼ増幅、逆PCR、ピコタイターPCR、およびエマルジョンPCRを含む。他の適切な増幅方法は、リガーゼ連鎖反応（LCR）、転写増幅、自家持続配列複製法、標的ポリヌクレオチド配列の選択的な増幅、コンセンサス配列プライムポリメラーゼ連鎖反応（CP-PCR）、任意プライムポリメラーゼ連鎖反応（AP-PCR）、縮重オリゴヌクレオチドプライムPCR（DOP-PCR）、および核酸ベース配列増幅（NABSA）を含む。本明細書で使用され得る他の増幅方法は、米国特許第5,242,794号、第5,494,810号、第4,988,617号、および第6,582,938号を参照。

【0442】

特定の実施形態において、PCRは、DNA分子が個々のパーティションへ分注された後にDNA分子を増幅するために使用される。場合によっては、増幅アダプター内の1つ以上の特異的なプライミング配列が、PCR増幅のために利用される。増幅アダプターは、個々のパーティションへの分注の前または後に、断片化DNA分子にライゲーションされ得る。両方の末端上で適切なプライミング配列を有する増幅アダプターを含むポリヌクレオチドは、指数関数的に増幅されたPCRであり得る。例えば、プライミング配列を含む増幅アダプターの不完全なライゲーション効率が原因で適切なプライミング配列を1つしか持たないポリヌクレオチドは、線形増幅のみを受けることもある。さらに、適切なプライミング配列を含むいかなるアダプターもライゲーションされない場合、ポリヌクレオチドは、増幅、例えば、PCR増幅からまとめて除去され得る。いくつかの実施形態において、PCRサイクルの数は10～30の間で変動するが、9、8、7、6、5、4、3、2、もしくはそれ以下にまで低くなるか、または40、45、50、55、60、またはそれ以上にまで高くなり得る。その結果、適切なプライミング配列を有する増幅アダプ

ターを運ぶ指数関数的に増幅可能な断片は、PCR増幅の後、線形に増幅可能または増幅可能ではない断片と比較して、はるかに高い（1000倍以上）濃度で存在し得る。全ゲノム増幅技術（例えば、ランダム化プライマーを用いる増幅、またはphi29ポリメラーゼを用いる多置換増幅）と比較したPCRの利点は、限定されないが、より均一な相対的な配列カバレッジ—各断片をサイクル毎に多くとも1回コピーすることができ、増幅がサーモサイクリングプログラムによって制御されるので—、例えば、MDAよりも実質的に低いキメラ分子を形成する速度（Lasken et al., 2007, BMC Biotechnology）—キメラ分子が、アセンブリグラフにおいて非生物学的配列を提示することによる正確な配列アセンブリに対する重大な問題を提起するため。これは、高い率のミスアセンブリ、または非常に曖昧で断片化されたアセンブリをもたらし—、特異的配列を有する特異的プライミング部位を使用することと比較して、MDAにおいて通常使用されるランダム化プライマーの結合から生じ得る配列特異的バイアスを減少させること、PCRサイクル数の選択によって制御することができる最終的な増幅DNA産物の量におけるより高い再現性、および通常的全ゲノム増幅技術と比較してPCRにおいて通常使用されるポリメラーゼによる複製におけるより高い忠実度を含む。

10

20

30

40

50

【0443】

いくつかの実施形態において、fill-in反応は、第1および第2のプライマーを使用した1つ以上の標的ポリヌクレオチドの増幅を伴うか、または上記増幅の一部として実行され、第1のプライマーは、第1のアダプターオリゴヌクレオチドの1つ以上の補体の少なくとも一部にハイブリダイズ可能な配列を含み、さらに、第2のプライマーは、第2のアダプターオリゴヌクレオチドの1つ以上の補体の少なくとも一部にハイブリダイズ可能な配列を含む。第1および第2のプライマーのそれぞれは、任意の適切な長さ、例えば、約10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100、もしくはそれ以上のヌクレオチド、上記よりも少ないヌクレオチド、または上記よりも多いヌクレオチドであり得、その一部またはすべては、対応する標的配列に相補的であり得る（例えば、約5、10、15、20、25、30、35、40、45、50、またはそれ以上のヌクレオチド、上記よりも少ないヌクレオチド、または上記よりも多いヌクレオチド）。例えば、約10～50のヌクレオチドが対応する標的配列に相補的であり得る。

【0444】

「増幅」は、標的配列のコピー数が増加する任意のプロセスを指す。場合によっては、複製反応は、ポリヌクレオチドの単一の相補的なコピー／レプリカのみを産生し得る。標的ポリヌクレオチドのプライマー指向性の増幅のための方法は、限定されないが、ポリメラーゼ連鎖反応（PCR）に基づく方法を含む。PCRによる標的配列の増幅に好ましい条件は、プロセスにおける様々な工程で最適化され、かつ、標的のタイプ、標的の濃度、増幅される配列の長さ、標的および／または1つ以上のプライマーの配列、プライマーの長さ、プライマーの濃度、使用するポリメラーゼ、反応量、1つ以上の要素対1つ以上の他の要素の比などの反応における要素の特徴に依存し、それらの一部またはすべては変更することができる。一般に、PCRは、（二本鎖の場合に）増幅される標的の変性、1つ以上のプライマーの標的へのハイブリダイゼーション、およびDNAポリメラーゼによるプライマーの伸長の工程を含み、工程は、標的配列を増幅するために繰り返される（または「サイクルされる」）。このプロセスにおける工程は、収量を増強するために、擬似的な産物の形成を減らすために、および／またはプライマーのアニールの特異性を増加または減少させるなどのために、様々な結果について最適化され得る。最適化の方法は、限定されないが、増幅反応における要素のタイプまたは数に対する、および／または、特定の工程での温度、特定の工程の持続時間、および／またはサイクルの数などのプロセスにおける所定の工程の条件に対する調整を含む。

【0445】

いくつかの実施形態において、増幅反応は少なくとも約5、10、15、20、25、30、35、40、50、60、70、80、90、100、150、200、またはそ

れ以上のサイクルを含み得る。いくつかの例において、増幅反応は、少なくとも約 20、25、30、35、または 40 のサイクルを含み得る。いくつかの実施形態において、増幅反応は、約 5、10、15、20、25、35、40、50、60、70、80、90、100、150、200、またはそれ以上のサイクル以下を含む。サイクルは、1、2、3、4、5、6、7、8、9、10、またはそれ以上の工程など、任意の数の工程を含み得る。工程は、限定されないが、3' 末端伸長（例えば、アダプター fill-in）、プライマーアニーリング、プライマー伸長、および鎖変性を含む所定の工程の目的を達成するのに適切な、任意の温度または温度の勾配を含むことができる。工程は、限定されないが、手動で中断されるまで無期限に含む、約 1、5、10、15、20、25、30、35、40、45、50、55、60、70、80、90、100、120、180、240、300、360、420、480、540、600、1200、1800 秒、またはそれ以上、上記の秒未満、または上記の秒を超える持続時間を含む、任意の持続時間であり得る。異なる工程を含む任意の数のサイクルは、任意の順で組み合わせられ得る。いくつかの実施形態において、異なる工程を含む異なるサイクルは、組み合わせるのサイクルの総数が約 5、10、15、20、25、30、35、40、50、60、70、80、90、100、150、または 200、またはそれ以上のサイクル、上記よりも少ないサイクル、あるいは上記よりも多いサイクルとなるように、組み合わせられる。いくつかの実施形態において、増幅は fill-in 反応の後に行なわれる。

10

【0446】

いくつかの実施形態において、増幅反応は、少なくとも約 1、2、3、4、5、6、7、8、9、10、12、14、16、18、20、25、30、40、50、100、200、300、400、500、600、800、1000 ng の標的 DNA 分子上で実行され得る。他の実施形態において、増幅反応は、約 1、2、3、4、5、6、7、8、9、10、12、14、16、18、20、25、30、40、50、100、200、300、400、500、600、800、1000 ng 未満の標的 DNA 分子上で実行され得る。

20

【0447】

増幅は、独立した試料から標的ポリヌクレオチドをプールする前または後に実行され得る。

【0448】

本開示の方法は、試料に存在する増幅可能な核酸の量を判定する工程を含む。任意の既知の方法は増幅可能な核酸を定量化するために使用され得、典型的な方法はポリメラーゼ連鎖反応（PCR）、具体的には定量的ポリメラーゼ連鎖反応（qPCR）である。qPCR はポリメラーゼ連鎖反応に基づく技術であり、標的とされた核酸分子を増幅し、かつ同時に定量化するために使用される。qPCR は、DNA 試料中の特異的配列の検出および定量化（DNA 入力または追加の正規化遺伝子へと正規化される際のコピーまたは相対量の絶対数として）の両方を可能にする。手順は、増幅された DNA が各増幅サイクルの後にリアルタイムでの反応において蓄積すると定量化される追加の特徴と共に、ポリメラーゼ連鎖反応の一般的な原理に従う。QPCR は、例えば、Kurnit ら（米国特許第 6,033,854 号）、Wang ら（米国特許第 5,567,583 号および第 5,348,853 号）、Ma ら（The Journal of American Science, 2(3), 2006）、Heid ら（Genome Research 986-994, 1996）、Sambrook および Russell（Quantitative PCR, Cold Spring Harbor Protocols, 2006）、および Higuchi（米国特許第 6,171,785 号および第 5,994,056 号）に記載されている。これらの内容は全体として本明細書での引用により組み込まれる。

30

40

【0449】

定量化の他の方法は、二本鎖 DNA と挿入する蛍光色素、および、相補的 DNA でハイブリダイズされた時に蛍光を発する修飾された DNA オリゴヌクレオチドプローブの使用

50

を含む。これらの方法は広く使用することができるが、具体的には一例としてさらに詳細に記載されるようなリアルタイムPCRにも適合する。第1の方法において、DNA結合色素はPCRにおいてすべての二本鎖(ds)DNAに結合し、その結果色素の蛍光が生じる。したがって、PCR中のDNA産物の増加は、蛍光強度の増加を引き起こし、各サイクルにて測定され、DNA濃度の定量化を可能にする。反応は、蛍光性の(ds)DNA色素の追加により、標準PCR反応と同様に調製される。反応はサーモサイクラー中で実行され、各サイクルの後、蛍光のレベルが検出器で比較され、(ds)DNA(すなわち、PCR産物)に結合している時に、色素は蛍光を発するのみである。標準希釈液に関して、PCRにおける(ds)DNA濃度が測定され得る。他のリアルタイムPCR方法のように、得られた値は、その値に関連する絶対単位を持たない。測定されたDNA/RNA試料と標準希釈液との比較は、標準希釈液に対する試料の割合または比率をもたらし、異なる組織または実験条件の間の相対的な比較を可能にする。標的遺伝子の定量化および/または発現における正確性を確保することは、安定して発現された遺伝子に関して標準化され得る。未知の遺伝子のコピー数は、既知のコピー数の遺伝子に対して同様に正規化され得る。

10

20

30

40

50

【0450】

第2の方法は、プローブ配列を含有するDNAのみを定量化するために、配列特異的なRNAまたはDNAベースのプローブを使用し、したがって、レポータープローブの使用は、特異性を著しく増大させ、いくつかの非特異的なDNA増幅の存在下でさえも定量化を可能にする。これにより、多重化、すなわち、異なるように色を付けられた標識を伴う特異的なプローブを使用することにより同じ反応における様々な遺伝子についてのアッセイを可能にするが、ただし、すべての遺伝子は同様の効率で増幅されるものとする。

【0451】

この方法は、プローブの1つの末端では蛍光レポーター(例えば、6-カルボキシフルオレセイン)、およびプローブの対向する末端ではクエンチャー(例えば、6-カルボキシテトラメチルローダミン)を有するDNAベースのプローブで一般的に実行される。クエンチャーに対するレポーターの近接性は、その蛍光の検出を妨げる。ポリメラーゼ(例えば、Taqポリメラーゼ)の5'~3'のエキソヌクレアーゼ活性によるプローブの破壊は、レポーター-クエンチャーの近接性を破壊し、したがって、蛍光のクエンチされていない発光を可能にし、これは検出することができる。各PCRサイクルにてレポータープローブにより標的とされる産物の増加の結果、プローブの破壊およびレポーターの放出に起因する蛍光の比例的な増加が生じる。反応は標準PCR反応と同様に調製され、レポータープローブが加えられる。反応が始まると、PCRのアニーリング段階中に、プローブとプライマーの両方がDNA標的にアニーリングされる。新しいDNA鎖の重合はプライマーから始まり、いったんポリメラーゼがプローブに到達すると、その5'-3'-エキソヌクレアーゼはプローブを分解し、クエンチャーから蛍光レポーターを物理的に分離して、結果として蛍光の増加をもたらす。蛍光はリアルタイムPCRサーモサイクラーにおいて検出かつ測定され、産物の指数関数的な増加に対応する蛍光の幾何学的な増加は、各反応における閾値サイクルを判定するために使用される。

【0452】

反応の対数期中に存在するDNAの相対濃度は、対数目盛上でサイクル数に対する蛍光をプロットすることにより判定される(そのため、指数関数的に増大する量は直線をもたらす)。バックグラウンド上での蛍光の検出のための閾値が判定される。試料からの蛍光が閾値を超えるサイクルは、サイクル閾値、C_tと呼ばれる。DNAの量は対数期中のすべてのサイクルで二倍になるため、DNAの相対量を計算することができ、例えば、他のものよりも3サイクル早いC_tを有する試料は、 $2^3 = 8$ 倍多い鋳型を有する。その後、核酸(例えば、RNAまたはDNA)の量は、結果を既知量の核酸の連続希釈(例えば、希釈なし、1:4、1:16、1:64)のリアルタイムPCRにより生成された標準曲線と比較することにより判定される。

【0453】

特定の実施形態において、qPCR反応は、蛍光共鳴エネルギー転移（FRET）、例えば、LIGHTCYCLERハイブリダイゼーションプローブを利用する二重フルオロフォア手法を含み、2つのオリゴヌクレオチドプローブはアンプリコンにアニーリングされる（例えば、米国特許第6,174,670号を参照）。オリゴヌクレオチドは、効率的なエネルギー転移に適合する距離で分離されたフルオロフォアでヘッド-テールの配向（head-to-tail orientation）においてハイブリダイズするように設計される。核酸に結合されたまたは伸長産物に組み込まれた時にシグナルを発するように構造化される標識されたオリゴヌクレオチドの他の例は、SCORPIONSプローブ（例えば、Whitcombe et al., Nature Biotechnology 17:804-807, 1999、および米国特許第6,326,145号）、Sunrise（またはAMPLIFLOUR）プライマー（例えば、Nazarenko et al., Nuc. Acids Res. 25:2516-2521, 1997、および米国特許第6,117,635号）、ならびにLUXプライマーおよびMOLECULAR BEACONSプローブ（例えば、Tyagi et al., Nature Biotechnology 14:303-308, 1996および米国特許第5,989,823号）を含む。

10

【0454】

他の実施形態において、qPCR反応は、蛍光Taqman方法、およびリアルタイムで蛍光を測定することが可能な機器（例えば、ABI Prism 7700 Sequence Detector）を使用する。Taqman反応は、2つの異なる蛍光色素で標識されるハイブリダイゼーションプローブを使用する。1つの色素はレポーター色素（6-カルボキシフルオレセイン）であり、もう1つはクエンチング色素（6-カルボキシテトラメチルローダミン）である。プローブが無傷の場合、蛍光エネルギー転移が生じ、レポーター色素の蛍光的な発光がクエンチング色素により吸収される。PCRサイクルの伸長段階中に、蛍光ハイブリダイゼーションプローブは、DNAポリメラーゼの5'-3'の核酸分解活性により切断される。プローブの切断時に、レポーター色素の発光はこれ以上クエンチング色素へと効率的に転移されず、その結果、レポーター色素の蛍光発光スペクトルが増加する。リアルタイムの方法または単点の検出方法を含む、任意の核酸定量化方法は、試料中の核酸の量を定量化するために使用され得る。検出は、様々な異なる方法（例えば、染色、標識されたプローブによるハイブリダイゼーション、ビオチン化プライマーの組み込みと、その後のアビジン-酵素コンジュゲートの検出、増幅されたセグメントへのdCTPまたはdATPなどの32P標識デオキシヌクレオチド三リン酸塩の組み込み）に加えて、核酸定量化のための任意の他の適切な検出方法によって実行され得る。定量化は増幅工程を含むこともあれば、含まないこともある。

20

30

【0455】

いくつかの実施形態において、本開示は、結合されたDNAセグメントを同定または定量化するための標識を提供する。場合によっては、連鎖されたDNAセグメントは、アレイのハイブリダイゼーションなどの下流での用途を補助するために標識され得る。例えば、連鎖されたDNAセグメントは、ランダムなプライミングまたはニックトランスレーションを使用して標識され得る。

40

【0456】

多種多様な標識（例えば、レポーター）は、限定されないが増幅工程中を含んで、本明細書に記載されるヌクレオチド配列を標識するために使用されてもよい。適切な標識は、放射性核種、酵素、蛍光、化学発光剤、または発色剤に加え、リガンド、補助因子、阻害剤、磁気微粒子などを含む。そのような標識の例は、米国特許第3,817,837号、米国特許第3,850,752号、米国特許第3,939,350号、米国特許第3,996,345号、米国特許第4,277,437号、米国特許第4,275,149号、および米国特許第4,366,241号に含まれており、これらは全体として引用により組み込まれる。

【0457】

50

追加の標識は、限定されないが、 β -ガラクトシダーゼ、インペルターゼ、緑色蛍光タンパク質、ルシフェラーゼ、クロラムフェニコール、アセチルトランスフェラーゼ、 β -グルクロニダーゼ、エキソ-グルカナーゼ (exo-glucanase)、およびグルコアミラーゼを含む。蛍光標識も、特定の化学的性質で特異的に合成された蛍光試薬と同様に使用されてもよい。蛍光を測定する多種多様な方法が利用可能である。例えば、いくつかの蛍光標識は、励起または発光のスペクトルの変化を示し、いくつかの蛍光標識は1つの蛍光レポーターが蛍光を失い、一方で第2の蛍光レポーターが蛍光を増加させる共鳴エネルギー転移を示し、いくつかの蛍光標識は蛍光の消失 (クエンチ) または出現を示し、いくつかの蛍光標識は回転動作を報告する。

【0458】

10

さらに、標識化のための十分な材料を得るために、1つの反応当たりの増幅サイクルの数を増やす代わりに、複数の増幅がプールされ得る。代替的に、標識されたヌクレオチドは、増幅反応の最後のサイクル、例えば、PCRの30のサイクル (標識なし) + PCRの10のサイクル (と標識) に組み込まれ得る。

【0459】

特定の実施形態において、本開示は、結合されたDNAセグメントに結合することができるプローブを提供する。本明細書で使用されるように、「プローブ」という用語は、関心対象の別の分子 (例えば、別のオリゴヌクレオチド) にハイブリダイズすることができる分子 (例えば、精製された制限消化物などにおいて自然に生じるかどうか、または合成的に、組換え的に、もしくはPCR増幅により産生されたかどうかにかかわらず、オリゴヌクレオチド) を指す。プローブは、オリゴヌクレオチドである場合、一本鎖または二本鎖であり得る。プローブは、特定の標的 (例えば、遺伝子配列) の検出、同定、および単離に役立つ。場合によっては、プローブは、限定されないが、酵素 (例えば、ELISAに加え、酵素に基づく組織化学アッセイ)、蛍光、放射性、および発光のシステムを含む任意の検出システムにおいて検出可能となるような標識に関連付けら得る。

20

【0460】

アレイおよびマイクロアレイに関して、「プローブ」という用語は、プローブへとハイブリダイズしたヌクレオチド配列を検出するためにアレイに固定される任意のハイブリダイズ可能な物質を指すために使用される。場合によっては、プローブは、約10bp~500bp、約10bp~250bp、約20bp~250bp、約25bp~200bp、約25bp~100bp、約30bp~100bp、または約30bp~80bpであり得る。場合によっては、プローブは、約10bp、約20bp、約30bp、約40bp、約50bp、約60bp、約70bp、約80bp、約90bp、約100bp、約150bp、約200bp、約250bp、約300bp、約400bp、または約500bp長であり得る。例えば、プローブは約20~約50bp長であり得る。プローブ設計の例および論理的根拠は、WO95/11995、EP717,113、およびWO97/29212で見つけることができる。

30

【0461】

プローブ、プローブのアレイ、またはプローブのセットは、支持体上で固定され得る。支持体 (例えば、固体支持体) はガラス、シリカ、プラスチック、ナイロン、またはニトロセルロースなどの様々な材料で作られ得る。支持体は剛性であり得、平らな表面を有し得る。支持体は、約1~10,000,000個の分離した遺伝子座 (resolved loci) を有し得る。例えば、支持体は、約10~10,000,000、約10~5,000,000、約100~5,000,000、約100~4,000,000、約1000~4,000,000、約1000~3,000,000、約10,000~3,000,000、約10,000~2,000,000、約100,000~2,000,000、または約100,000~1,000,000の分離した遺伝子座を有し得る。分離した遺伝子座の密度は、1平方センチメートル内で少なくとも約10、約100、約1000、約10,000、約100,000、または約1,000,000の分離した遺伝子座であり得る。場合によっては、分離した遺伝子座の各々は、95%を超える

40

50

単一のタイプのオリゴヌクレオチドによって占有され得る。他の場合、分離した遺伝子座の各々は、プローブのプールされた混合物またはプローブのセットにより占有され得る。さらなる場合において、いくつかの分離した遺伝子座は、プローブのプールされた混合物またはプローブのセットにより占有され、他の分離した遺伝子座は95%を超える単一のタイプのオリゴヌクレオチドにより占有される。

【0462】

場合によっては、アレイ上の所定のヌクレオチド配列のためのプローブの数は、そのようなアレイにハイブリダイズされるDNA試料に対し非常に過剰な場合がある。例えば、アレイは、入力された試料におけるDNAの量に対して約10、約100、約1000、約10,000、約100,000、約1,000,000、または約100,000、000倍のプローブ数を有し得る。

10

【0463】

場合によっては、アレイは、約10、約100、約1000、約10,000、約100,000、約1,000,000、約10,000,000、または約1,000,000,000のプローブを有し得る。

【0464】

プローブのアレイまたはセットは、支持体上で段階ごとに合成され得るか、またはあらかじめ合成された形態で結合され得る。合成方法の1つは、高密度の小型化したアレイにおけるオリゴヌクレオチドプローブの合成を指示するための光の使用を伴う、VLSIPS（商標）（米国特許第5,143,854号およびEP476,014に記載される）である。合成サイクルの数を減らすためのマスクの設計のためのアルゴリズムは、米国特許第5,571,639号および米国特許第5,593,839号に記載されている。アレイはさらに、EP624,059に記載されるように、機械的に制限された流路により支持体の細胞にモノマーを送達することにより、組み合わせて合成され得る。アレイはさらに、インクジェットプリンターを使用して支持体上へと試薬をスポットする（spotting）ことにより合成され得る（例えば、EP728,520を参照）。

20

【0465】

いくつかの実施形態において、本開示は、アレイ上に連鎖されたDNAセグメントをハイブリダイズする方法を提供する。「基材」または「アレイ」は、合成的または生合成的のいずれかで調製され得、様々な異なるフォーマット（例えば、可溶性分子のライブラリー、および、樹脂ビーズ、シリカチップ、または他の固体支持体にテザリングされたオリゴヌクレオチドのライブラリー）において生物活性についてスクリーニングされ得る核酸の故意に作り出された集合体である。加えて、「アレイ」という用語は、基材上へと実質的に任意の長さ（例えば、1～約1000長のヌクレオチドモノマー）の核酸をスポットすることにより調製され得る核酸のそのようなライブラリーを含んでいる。

30

【0466】

アレイの技術、ならびに様々な関連技術およびその適用は、一般的に多くのテキストブックおよび文書で記載されている。例えば、これらの文書として、Lemieux et al., 1998, Molecular Breeding 4, 277-289; Schena and Davis, Parallel Analysis with Biological Chips. in PCR Methods Manual (eds. M. Innis, D. Gelfand, J. Sninsky); Schena and Davis, 1999, Genes, Genomes and Chips. In DNA Microarrays: A Practical Approach (ed. M. Schena), Oxford University Press, Oxford, UK, 1999; The Chipping Forecast (Nature Genetics special issue; January 1999 Supplement); Mark Schena (Ed.), Microarray Biochip Technology, (Eaton Publishing Company); Cortes, 2

40

50

000, The Scientist 14 [17]:25; Gwynn and Page, Microarray analysis: the next revolution in molecular biology, Science, 1999 Aug. 6、および、Eakins and Chu, 1999, Trends in Biotechnology, 17, 217-218が挙げられる。

【0467】

通常、任意のライブラリーは、ライブラリーのメンバーを空間的に分離することにより、アレイへと順序正しく配置され得る。配列に適切なライブラリーの例は、核酸ライブラリー（DNA、cDNA、オリゴヌクレオチドなどのライブラリーを含む）、ペプチド、ポリペプチド、およびタンパク質のライブラリー、ならびに、任意の分子を含むライブラリー、例えば、とりわけ、リガンドライブラリーを含む。

10

【0468】

ライブラリーは、メンバーの拡散および混合を制限するために、固相（例えば、固体の基材）上へ固定化または固定され得る。場合によっては、DNA結合リガンドのライブラリーが調製され得る。とりわけ、ライブラリーは、膜および非多孔基材、例えば、プラスチックやガラスを含む実質的に平らな固相に固定され得る。さらに、ライブラリーは、インデキシング（すなわち、特定のメンバーへの言及またはアクセス）が容易になるような方法で配置され得る。いくつかの例において、ライブラリーのメンバーは、格子形態のスポットとして適用され得る。一般的なアッセイ系はこの目的に適合し得る。例えば、アレイは、ウェルに複数のメンバーを有するか、または各ウェルに1つのメンバーを有するマイクロプレートの表面上で固定され得る。さらに、固体の基材は、ニトロセルロースまたはナイロンの膜（例えば、プロッティング実験に使用される膜）などの膜でもよい。代替的な基材は、ガラスまたはシリカベースの基材を含む。したがって、ライブラリーは、任意の適切な方法、例えば、電荷相互作用、あるいは、ウェルの壁もしくは底、または膜の表面への化学結合により、固定され得る。他の配置および固定化の手段、例えば、ピペット操作、ドロップタッチ、圧電手段、インクジェットおよびバブルジェット技術、静電気の適用などが使用され得るシリコンベースのチップの場合、フォトリソグラフィーがチップ上でライブラリーを配置および固定化するために利用され得る。

20

【0469】

ライブラリーは、固体基材上に「スポットされる（spotted）」ことにより配置され、これは、手動で、またはメンバーを堆積させるためにロボット工学を利用することにより行われてもよい。一般に、アレイはマクロアレイまたはマイクロアレイと記載されてもよく、違いはスポットの大きさである。マクロアレイは、約300ミクロン以上のスポットサイズを含み得、既存のゲルおよびプロットスキャナーにより容易に画像化され得る。マイクロアレイにおけるスポットサイズは、直径200ミクロン未満であり得、これらのアレイは通常、数千ものスポットを含んでいる。したがって、マイクロアレイは、特別なロボット工学および画像化機器を必要とし得、これらは、カスタム作製される必要があり得る。器具は概して、Cortese, 2000, The Scientist 14 [11]:26による総説に記載されている。

30

【0470】

DNA分子の固定されたライブラリーを作製するための技術は記載されている。通常、このような方法のほとんどは、例えば、固体の基材上の様々な別個の位置で配列の様々な順列を構築するためにマスキング技術を使用して、一本鎖核酸分子ライブラリーを合成する方法について記載している。米国特許第5,837,832号は、超大規模集積回路技術に基づいてシリコン基板に固定されたDNAアレイを作製するための改善された方法について記載している。特に、米国特許第5,837,832号は、本開示の固定されたDNAライブラリーを作製するために使用され得る基材上の空間的に定められた場所でプローブの特定のセットを合成するために「タイリング」と呼ばれる戦略について記載している。米国特許第5,837,832号はさらに、使用され得る初期の技術に対する言及を提供する。他の場合では、アレイはさらに、光析出化学（photo deposite

40

50

on chemistry) を使用して構築され得る。

【0471】

ペプチド（またはペプチド模倣体）のアレイも、アレイにおける別個のあらかじめ定められた場所で別個のライブラリーのメンバー（例えば、固有のペプチド配列）それぞれを配する様式で、表面上で合成され得る。ライブラリーのメンバーそれぞれの同一性は、アレイにおけるその空間的位置により判定される。あらかじめ定められた分子（例えば、標的またはプローブ）と反応的なライブラリーのメンバーとの間の結合相互作用が生じるアレイにおける位置が判定され、それにより、空間的位置に基づき反応的なライブラリーのメンバーの配列が同定される。これらの方法は、米国特許第5,143,854号、WO 90/15070、およびWO 92/10092、Fodor et al. (1991) Science, 251:767; Dower and Fodor (1991) Ann. Rep. Med. Chem., 26:271に記載されている。

10

【0472】

検出を支援するために、任意の容易に検出可能なレポーター、例えば、蛍光性、生物発光性、リン光性、放射性などのレポーターといった標識を使用することができる（上記で議論されるように）。そのようなレポーター、それらの検出、標的／プローブへの結合などは、本文書の他の場所で議論される。プローブおよび標的の標識化はさらに、Shalton et al., 1996, Genome Res 6(7):639-45に開示されている。

20

【0473】

いくつかの市販で入手可能なマイクロアレイのフォーマットの例は、Marshall and Hodgson, 1998, Nature Biotechnology, 16(1), 27-31に説明されている。

【0474】

アレイベースのアッセイからデータを生成するために、シグナルは、プローブとヌクレオチド配列との間のハイブリダイゼーションの存在または不在を示すために検出され得る。さらに、直接および間接的な標識化技術も利用され得る。例えば、直接的な標識化は、アレイに関連するプローブにハイブリダイズするヌクレオチド配列へ直接的に蛍光色素を組み込む（例えば、色素は、標識化ヌクレオチドまたはPCRプライマーの存在下で酵素的合成によりヌクレオチド配列に組み込まれる）。直接的な標識化スキームは、例えば、同様の化学的な構造および特徴を有する蛍光色素のファミリーの使用により、強固なハイブリダイゼーションシグナルをもたらす、かつ簡単に実施することができる。核酸の直接的な標識化を含む場合には、シアニンまたはアレクサ (alex a) のアナログが、複数の蛍光比較アレイ解析 (multiple-fluor comparative array analyses) に利用され得る。他の実施形態において、間接的な標識化スキームは、マイクロアレイプローブへのハイブリダイゼーションの前または後に、核酸にエピトープを組み込むために利用され得る。1つ以上の染色の手順および試薬は、ハイブリダイズされた複合体（例えば、エピトープに結合し、それによりハイブリダイズされた種のエピトープへの色素分子のコンジュゲーションによって蛍光シグナルをもたらす、蛍光分子）を標識するために使用され得る。

30

40

【0475】

配列決定

様々な実施形態において、本明細書に記載されるか、そうでなければ既知の適切な配列決定方法が、試料内の核酸分子から配列情報を得るために使用される。配列決定は、古典的なサンガー配列決定方法を介して達成することができる。配列はさらに、ハイスループットシステムを使用して達成され、その一部は、成長鎖への組み込みの直後またはその際の配列決定されたヌクレオチドの検出、すなわち、リアルタイムまたはほぼリアルタイムでの配列の検出を可能にする。場合によっては、ハイスループット配列決定は、1時間につき少なくとも1,000、少なくとも5,000、少なくとも10,000、少なくとも20,000、少なくとも30,000、少なくとも40,000、少なくとも50,

50

000、少なくとも100、000、または少なくとも500、000の配列リードを生成し、ここで、配列決定リードは、1つのリードにつき少なくとも約50、約60、約70、約80、約90、約100、約120、約150、約180、約210、約240、約270、約300、約350、約400、約450、約500、約600、約700、約800、約900、または約1000の塩基であり得る。

【0476】

配列決定は、関心対象の特定の領域の濃縮の有無にかかわらず、全ゲノムであり得る。配列決定はゲノムの特定の領域を標的とすることができる。濃縮または標的化され得るゲノムの領域としては、限定されないが、単一遺伝子（またはその領域）、遺伝子パネル、遺伝子融合物、ヒト白血球抗原（HLA）遺伝子座（例えば、クラスI HLA-A、B、およびC、クラスII HLA-DRB1/3/4/5、HLA-DQA1、HLA-DQB1、HLA-DPA1、HLA-DPB1）、エクソン領域、エクソーム、および他の遺伝子座が挙げられる。ゲノム領域は、免疫応答、免疫レパートリー、免疫細胞多様性、転写（例えば、エキソーム）、癌（例えば、BRCA1、BRCA2、遺伝子のパネルまたはその領域、例えば、ホットスポット領域、体細胞バリエーション、SNV、増幅、融合、腫瘍突然変異負荷（TMB）、マイクロサテライト不安定性（MSI））、心疾患、遺伝性疾患、および他の疾患または状態に関連し得る。限定されないが配列捕捉を含む様々な方法を使用して、関心対象の領域を濃縮または標的化することができる。場合によっては、ライブラリー調製の前または後に配列捕捉工程（例えば、標的濃縮アレイによる）を用いて、捕捉Hi-C（CHi-C）またはCHi-C様プロトコルが使用される。

【0477】

いくつかの実施形態において、ハイスループット配列決定は、Illumina's Genome Analyzer IIX、MiSeqパーソナルシーケンサー、またはHiSeqシステム、例えば、HiSeq 2500、HiSeq 1500、HiSeq 2000、またはHiSeq 1000の機械を使用するものなどにより利用可能な技術の使用を含む。これらの機械は、合成化学による可逆的なターミネーターベースの配列決定を使用する。これらの機械は、8日で2000億以上のDNAのリードを行うことができる。3日、2日、1日以下の時間内で実行するために、より小さなシステムが利用されてもよい。

【0478】

いくつかの実施形態において、ハイスループット配列決定は、ABI Solid Systemにより利用可能な技術の使用を含む。クローン増幅されたDNA断片の超並列配列決定を可能にするこの遺伝子解析プラットフォームは、ビーズに結合した。配列決定の方法論は、色素で標識したオリゴヌクレオチドによる連続的なライゲーションに基づく。

【0479】

次世代配列決定は、（例えば、Life Technologies（Ion Torrent）の技術を使用する）イオン半導体配列決定を含み得る。イオン半導体配列決定は、ヌクレオチドがDNAの鎖へ組み込まれる時にイオンが放出され得るという事実を利用することができる。イオン半導体配列決定を行なうために、微細加工されたウェルの高密度アレイを形成することができる。各ウェルは単一のDNA鋳型を保持することができる。ウェルの下はイオン感受性層であり、イオン感受性層の下はイオンセンサーであり得る。ヌクレオチドがDNAに加えられると、H⁺が放出され、pHの変化として測定され得る。H⁺イオンは電圧に変換され、半導体センサーにより記録され得る。アレイチップをヌクレオチドで次々にあふれさせる（flooded）ことができる。走査、光、またはカメラは必要ではない。場合によっては、IONPROTON（商標）シーケンサーが核酸を配列決定するために使用される。場合によっては、IONPGM（商標）シーケンサーが使用される。Ion Torrent Personal Genome Machine（PGM）。PGMは2時間で1000万のリードを行うことができる。

【0480】

いくつかの実施形態において、ハイスループット配列決定は、Helicos Bio Sciences Corporation (Cambridge, Massachusetts) により利用可能な技術、例えば、合成による単一分子配列決定 (SMS S) 方法の使用を含む。SMS Sは、最大24時間で全ヒトゲノムの配列決定を可能にすることから、固有なものである。最終的に、SMS Sは、米国特許出願公開第20060024711号、第20060024678号、第20060012793号、第20060012784号、および第20050100932号に部分的に記載されている。

【0481】

いくつかの実施形態において、ハイスループット配列決定は、PicoTiterPlate装置などの454 Lifesciences, Inc. (Branford, Connecticut) により利用可能な技術の使用を含み、これは、機器においてCCDカメラにより記録される配列決定反応により生成される化学発光シグナルを伝達するファイバオプティックプレートを含む。このファイバオプティクスの使用は、4.5時間で最低2000万の塩基対の検出を可能にする。

【0482】

ビーズ増幅とその後のファイバオプティクスの検出を使用する方法は、Marguiles, M., et al. "Genome sequencing in microfabricated high-density picolitre reactors," Nature, doi:10.1038/nature03959、および、米国特許出願公開第20020012930号、第20030068629号、第20030100102号、第20030148344号、第20040248161号、第20050079510号、第20050124022号、および、第20060078909号に記載されている。

【0483】

いくつかの実施形態において、ハイスループット配列決定は、Clonal Single Molecule Array (Solexa, Inc.)、または可逆的なターミネーター化学を利用する合成による配列決定 (SBS) を使用して行なわれる。これらの技術は、米国特許第6,969,488号、第6,897,023号、第6,833,246号、第6,787,308号、および、米国特許出願公開第20040106110号、第20030064398号、第20030022207号、およびConstans, A., The Scientist 2003, 17(13):36に部分的に記載されている。

【0484】

次世代配列決定技術は、Pacific Biosciencesによるリアルタイム (SMRT (商標)) 技術を含み得る。SMRTにおいて、4つのDNA塩基の各々は、4つの異なる蛍光色素の1つに結合され得る。これらの色素はホスホ結合 (phospho linked) され得る。単一のDNAポリメラーゼは、ゼロモード導波路 (ZMW) の底部にある鑄型の一本鎖DNAの単一分子で固定され得る。ZMWは、(マイクロ秒で) ZMWの中および外で急速に拡散することができる蛍光ヌクレオチドのバックグラウンドに対する、DNAポリメラーゼによる単一のヌクレオチドの組み込みの観察を可能にする、閉じ込め構造 (confinement structure) となり得る。成長鎖にヌクレオチドを組み込むのに数ミリ秒かかる場合がある。この間に蛍光標識は励起され、蛍光シグナルを生成することができ、蛍光タグは切断され得る。ZMWは下から照らすことができる。励起ビームからの減衰光は、各ZMWの下部20~30nmを透過することができる。20zeptoliter (20×10^{-21} リットル) の検出限界を有する顕微鏡を作成することができる。小さな検出体積は、バックグラウンドノイズの低減において1000倍の改善を提供することができる。色素の対応する蛍光の検出は、どの塩基が組み込まれたかを示すことができる。このプロセスを繰り返すことができる。このプロセスは繰り返すことができる。

【0485】

10

20

30

40

50

場合によっては、次世代配列決定はナノ細孔配列決定である（例えば、Soni G V and Meller A. (2007) Clin Chem 53: 1996-2001を参照）。ナノ細孔は、直径約1ナノメートルの小さな穴であり得る。導電性流体におけるナノ細孔の浸漬、およびそこでの電位の適用は結果として、ナノ細孔を介したイオンの伝導が原因で、僅かな電流をもたらす得る。流れる電流の量はナノ細孔のサイズに感受性があり得る。DNA分子がナノ細孔を通ると、DNA分子上のヌクレオチドはそれぞれ、異なる程度にまでナノ細孔を塞ぐことができる。したがって、DNA分子がナノ細孔を通る際の、ナノ細孔を通る電流の変化は、DNA配列の読み取りを表すことができる。ナノ細孔配列決定技術は、Oxford Nanopore Technologies、例えば、GridIONシステムからのものであり得る。単一のナノ細孔は、マイクロウェルの上部にわたる高分子膜に挿入され得る。マイクロウェルはそれぞれ、個々の感知のための電極を有し得る。マイクロウェルは、1つのチップ当たり100,000以上のマイクロウェル（例えば、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、または1,000,000を超える）を有するアレイチップへと組み立てられ得る。機器（またはノード）がチップを解析するために使用され得る。データはリアルタイムで解析され得る。1つ以上の機器は一度に操作され得る。ナノ細孔は、タンパク質ナノ細孔、例えば、タンパク質アルファ溶血素、七量体タンパク質細孔であり得る。ナノ細孔は、ソリッドステートナノ細孔で出来ており、例えば、合成膜（例えば、SiNx、またはSiO₂）に形成されるナノメートルサイズの穴であり得る。ナノ細孔は、ハイブリッド細孔（例えば、ソリッドステート膜へのタンパク質細孔の統合）であり得る。ナノ細孔は、集積化したセンサーを有するナノ細孔であり得る（例えば、トンネル電極検出器、容量検出器、またはグラフェンベースのナノギャップもしくはエッジ状態検出器（例えば、Gara j et al. (2010) Nature vol. 67, doi: 10.1038/nature09379を参照））。ナノ細孔は、特定の型の分子（例えば、DNA、RNA、またはタンパク質）を解析するために官能化することができる。ナノ細孔配列決定は、「鎖配列決定」を含むことができ、ここで、無傷なDNAポリマーは、DNAが細孔を転移させる際にリアルタイムで配列決定しながらタンパク質ナノ細孔を通過することができる。酵素は二本鎖DNAの鎖を分離することができ、ナノ細孔を通じて鎖を与えることができる。DNAは一方の末端でヘアピンを有することができ、システムは両方の鎖を読み取ることができる。場合によっては、ナノ細孔配列決定は、「エキソヌクレアーゼ配列決定」であり、ここで、個々のヌクレオチドは前進性エキソヌクレアーゼによってDNA鎖から切断されることができ、そのヌクレオチドはタンパク質ナノ細孔を通過することができる。ヌクレオチドは、細孔内の分子（例えば、シクロデキストラン）に一過的に結合することができる。電流の特性的断絶を使用して、塩基を同定することができる。

10

20

30

40

50

【0486】

GENIAのナノ細孔配列決定技術を使用できる。操作したタンパク質細孔を脂質二重層の膜に包埋することができる。「能動的制御」技術を使用して、効率的なナノ細孔-膜アセンブリ、およびチャネルを通るDNA運動の制御を可能にする。場合によっては、ナノ細孔配列決定技術は、NABsys製である。ゲノムDNAは、平均長約100kbの鎖に断片化され得る。100kbの断片を一本鎖にし、その後、六量体のプローブでハイブリダイズさせることができる。プローブを有するゲノム断片は、ナノ細孔を通過して通り抜けることができ、電流対時間の追跡を作り出すことができる。電流追跡は、各ゲノム断片上のプローブの位置を提供することができる。ゲノム断片を並べて、ゲノムに対するプローブマップを作り出すことができる。そのプロセスは、プローブのライブラリーに対して並列に行うことができる。各プローブに対するゲノム長のプローブマップを生成することができる。誤りは、「移動窓配列決定バイハイブリダイゼーション (Sequencing By Hybridization) (mwSBH)」と称されるプロセスで修正することができる。場合によっては、ナノ細孔配列決定技術は、IBM/Roche製である。電子ビームを使用して、マイクロチップにナノ細孔サイズの開口部を作製すること

ができる。電界を使用して、ナノ細孔を通してDNAを引き寄せるまたはねじ込むことができる。ナノ細孔におけるDNAトランジスタ装置は、金属と誘電体が交互になったナノメートルサイズの層を含むことができる。DNA骨格中の別個の電荷を、電界によってDNAナノ細孔の内部に閉じ込めることができる。ゲート電圧をオフ／オンすることにより、DNA配列を読み取ることができる。

【0487】

次世代配列決定は、DNAナノボール配列決定を含むことができる（例えば、Complete Genomicsによって実施されるように。例えば、Drmanac et al. (2010) Science 327: 78-81を参照）。DNAを単離し、断片化し、サイズ選択することができる。例えば、DNAは（例えば、超音波処理によって）約500bpの平均長へと断片化することができる。アダプター（Ad1）を、断片の末端に結合させることができる。アダプターを使用して、配列決定反応のためのアンカーにハイブリダイズさせることができる。各末端に結合したアダプターを有するDNAを、PCR増幅することができる。アダプター配列は、相補的な一本鎖末端が互いに結合して環状DNAを形成するように、修飾することができる。DNAをメチル化して、その後の工程において使用されるIIS型制限酵素による切断からこれを保護することができる。アダプター（例えば、右のアダプター）は、制限認識部位を有することができ、制限認識部位はメチル化されないままであり得る。アダプターにおける非メチル化制限認識部位は、制限酵素（例えば、AclI）によって認識され得、DNAは、AclIによって右アダプターの右側に13bp切断されて、線形二本鎖DNAを形成することができる。第2のラウンドの右および左のアダプター（Ad2）を、線状DNAのいずれかの末端にライゲーションすることができ、両方のアダプターが結合しているDNAはすべて（例えば、PCRによって）PCR増幅することができる。Ad2配列を修飾して、それらが互いに結合して環状DNAを形成することを可能にする。DNAはメチル化することができるが、制限酵素認識部位は、左Ad1アダプターにおいてメチル化されないままであり得る。制限酵素（例えば、AclI）を適用することができ、DNAは、Ad1の左側13bpで切断され、線状DNA断片を形成することができる。第3ラウンドの右および左のアダプター（Ad3）を、線状DNAの右面および左面にライゲーションすることができ、その結果生じる断片をPCR増幅することができる。アダプターは、互いに結合して環状DNAを形成するように、修飾することができる。III型制限酵素（例えば、EcoP15）を加えることができ、EcoP15は、DNAをAd3の左に26bpおよびAd2の右に26bp切断することができる。この切断は、DNAの大きなセグメントを除去し、DNAを再び線状化することができる。第4ラウンドの右および左のアダプター（Ad4）をDNAにライゲーションすることができ、DNAを（例えば、PCRによって）増幅し、互いに結合して完成した環状DNA鑄型を形成するように、修飾することができる。

【0488】

ローリングサークル複製（例えば、Phi29 DNAポリメラーゼを使用する）を使用して、DNAの小さな断片を増幅することができる。4つのアダプター配列は、ハイブリダイズすることができるパンドローム配列を含有することができ、一本鎖をそれ自体の上に折りたたんで、平均で直径およそ200～300ナノメートルとなり得るDNAナノボール（DNB（商標））を形成することができる。DNAナノボールは、（例えば、吸着により）マイクロアレイ（配列決定フローセル）に結合させることができる。フローセルは二酸化ケイ素、チタン、およびヘキサメチルジシラザン（HMDS）、ならびにフォトレジスト材料でコーティングされたシリコンウエハであり得る。配列決定は、DNAに蛍光プローブをライゲーションすることによる未連鎖の配列決定によって実施することができる。調べた位置の蛍光の色は高解像度カメラによって可視化することができる。アダプター配列間のヌクレオチド配列の同一性を決定することができる。

【0489】

いくつかの実施形態において、ハイスループット配列決定は、AnyDot.chip

s (Genovox, Germany) を使用して行うことができる。特に、AnyDot.chips は、ヌクレオチド蛍光シグナル検出を 10 倍～50 倍増強できる。AnyDot.chips およびそれを使用する方法は、国際公開 WO 02088382、WO 03020968、WO 03031947、WO 2005044836、PCT/EP05/05657、PCT/EP05/05655、およびドイツ特許出願第 DE 101 49 786、DE 102 14 395、DE 103 56 837、DE 10 2004 009 704、DE 10 2004 025 696、DE 10 2004 025 746、DE 10 2004 025 694、DE 10 2004 025 695、DE 10 2004 025 744、DE 10 2004 025 745、および DE 10 2005 012 301 にある程度記載されている。

10

【0490】

他のハイスループット配列決定システムとしては、Venter, J., et al. Science 16 February 2001; Adams, M. et al. Science 24 March 2000、および、M. J. Levene et al. Science 299: 682-686, January 2003、ならびに、米国特許出願公開第 20030044781 号および第 2006/0078937 号に記載されるものが挙げられる。そのようなシステム全体は、核酸分子上で測定される重合反応による塩基の経時的な付加によって複数の塩基を有する標的核酸分子を配列決定することを含み、すなわち、配列決定される鋳型核酸分子上の核酸重合酵素の活性がリアルタイムで追跡される。次いで、塩基付加の配列における各工程での核酸重合酵素の触媒活性によって、どの塩基が標的核酸の成長中の相補鎖に組み込まれているかを同定することによって、配列を推定することができる。標的核酸分子複合体上のポリメラーゼは、標的核酸分子に沿って移動し、活性部位でオリゴヌクレオチドプライマーを伸長するのに適した位置に提供される。複数の標識型のヌクレオチドアナログが活性部位の近位に提供され、ヌクレオチドアナログの各識別可能な型は標的核酸配列中の異なるヌクレオチドに対して相補的である。成長核酸鎖は、ポリメラーゼを使用して活性部位で核酸鎖にヌクレオチドアナログを付加することによって伸長され、付加されるヌクレオチドアナログは、活性部位で標的核酸のヌクレオチドに対して相補的である。重合工程の結果としてオリゴヌクレオチドプライマーに付加されたヌクレオチドアナログが同定される。標識したヌクレオチドアナログを提供する工程と、成長核酸鎖を重合させる工程と、付加されたヌクレオチドアナログを同定する工程は繰り返され、それにより核酸鎖がさらに伸長され、標的核酸の配列が決定される。

20

30

【0491】

キット

特定の実施形態において、本開示は、本開示の 1 つ以上の構成要素を含むキットをさらに提供する。キットは、限定されないが、上に記載されたものを含む任意の適切な用途に使用することができる。キットは、例えば、複数の会合分子、固定剤、ヌクレアーゼ、リガーゼ、および／またはそれらの組み合わせを含むことができる。場合によっては、会合分子は、例えば、ヒストンを含むタンパク質であり得る。場合によっては、固定剤は、ホルムアルデヒド、または DSG、EGS、もしくは DSS を含む任意の他の DNA クロスリンキング剤であり得る。

40

【0492】

場合によっては、キットはさらに複数のビーズを含むことができる。ビーズは常磁性であり、および／または捕捉剤でコーティングされている。例えば、ビーズは、ストレプトアビジンおよび／または抗体でコーティングすることができる。

【0493】

場合によっては、キットは、アダプターオリゴヌクレオチドおよび／または配列決定プライマーを含むことができる。さらに、キットは、アダプターオリゴヌクレオチドおよび／または配列決定プライマーを使用してリード対を増幅することができる装置を含むこと

50

ができる。

【0494】

場合によっては、キットは、限定されないが、溶解緩衝液、ライゲーション試薬（例えば、dNTP、ポリメラーゼ、ポリヌクレオチドキナーゼ、および／またはリガーゼ緩衝液など）、およびPCR試薬（例えば、dNTP、ポリメラーゼ、および／またはPCR緩衝液など）を含むことができる。

【0495】

キットはさらに、キットの構成要素を使用するためのおよび／またはリード対を生成するための指示書を含むことができる。

【0496】

コンピュータおよびシステム

図3に示されるコンピュータシステム(500)は、媒体(511)および／またはネットワークポート(505)(固定化された媒体(512)を有するサーバー(509)に任意選択で接続され得る)から命令を読み取ることができるロジック装置として理解され得る。図3に示されるものなどのシステムは、CPU(501)、ディスクドライブ(503)、キーボード(515)および／またはマウス(516)などの任意選択の入力装置、ならびに任意選択のモニター(507)を含み得る。データ通信は指示された通信媒体を介して局所位置または遠隔位置のサーバーまで達成され得る。通信媒体は、データを送信および／または受信する任意の手段を含むことができる。例えば、通信媒体は、ネットワーク接続、無線接続、またはインターネット接続であってもよい。そのような接続は、ワールド・ワイド・ウェブ(World Wide Web)での通信を提供することができる。本開示に関するデータは、図3に例示されるように当事者(522)が受理および／または検討するために、そのようなネットワークまたは接続によって伝達され得ることが想定される。

【0497】

図4は、本開示の例示的な実施形態と関連して使用可能なコンピュータシステム(100)の第1の例のアーキテクチャを示すブロック図である。図4に描かれるように、例示的なコンピュータシステムは、命令を処理するためのプロセッサ(102)を含み得る。プロセッサの非限定的な例は、Intel Xeon(商標)プロセッサ、AMD Opteron(商標)プロセッサ、Samsung 32-bit RISC ARM 1176JZ(F)-Sv1.0(商標)プロセッサ、ARM Cortex-A8 Samsung S5PC100(商標)プロセッサ、ARM Cortex-A8 Apple A4(商標)プロセッサ、Marvell PXA 930(商標)プロセッサ、または機能的に同等のプロセッサを含む。実行の複数のスレッドが並列処理に使用可能である。いくつかの実施形態では、単一コンピュータシステム、クラスターであろうと、または複数のコンピュータ、携帯電話、および／またはパーソナルデータアシスタントデバイスを含むネットワーク上でシステムにわたって分配されようと、複数のプロセッサまたは複数のコアを有するプロセッサを使用することができる。

【0498】

図4に例示されるように、高速キャッシュ(104)は、プロセッサ(102)に接続するか、または組み込まれることで、プロセッサ(102)により近年使用されてきたまたは頻繁に使用されている命令またはデータのための高速メモリを提供することができる。プロセッサ(102)は、プロセッサバス(108)によりノースブリッジ(106)に接続される。ノースブリッジ(106)は、メモリバス(112)によりランダムアクセスメモリ(RAM)(110)に接続され、プロセッサ(102)によりRAM(110)へのアクセスを管理する。ノースブリッジ(106)はさらに、チップセットバス(116)によりサウスブリッジ(114)に接続される。サウスブリッジ(114)は次に、周辺バス(118)に接続される。周辺バスは、例えば、PCI、PCI-X、PCI Express、または他の周辺バスであってもよい。ノースブリッジおよびサウスブリッジはしばしば、プロセッサチップセットと称され、周辺バス(118)上でプロセ

10

20

30

40

50

ッサとRAMと周辺コンポーネントとの間のデータ転送を管理する。いくつかの代替的なアーキテクチャでは、ノースブリッジの機能性は、別のノースブリッジチップを使用する代わりにプロセッサに組み込まれ得る。

【0499】

いくつかの実施形態において、システム(100)は、周辺バス(118)に取り付けられるアクセラレータカード(122)を含み得る。アクセラレータは、フィールドプログラマブルゲートアレイ(FPGA)、または特定の処理を促進するための他のハードウェアを含んでもよい。例えば、アクセラレータは、適応データの再構成のために、または拡張された設定処理に使用される代数式を評価するために使用され得る。

【0500】

ソフトウェアとデータは、外部ストレージ(124)に記憶され、プロセッサにより使用されるRAM(110)および／またはキャッシュ(104)へとロードすることができる。システム(100)は、システムリソースの管理のためのオペレーティングシステムを含み、オペレーティングシステムの非限定的な例は、Linux(登録商標)、Windows(商標)、MACOS(商標)、BlackBerry OS(商標)、iOS(商標)、および他の機能的に同等なオペレーティングシステムの他に、本開示の例示的な実施形態に従ってデータの記憶および最適化を管理するためのオペレーティングシステム上で実行されるアプリケーションソフトウェアを含む。

【0501】

この例において、システム(100)はさらに、ネットワーク接続ストレージ(NAS)などの外部記憶装置、および分散並列処理に使用され得る他のコンピュータシステムに対するネットワークインターフェースを提供するために、周辺バスに接続されるネットワークインターフェースカード(NIC)(120)および(121)を含む。

【0502】

図5は、複数のコンピュータシステム(202a)および(202b)、複数の携帯電話およびパーソナルデータアシスタント(202c)、ならびにネットワーク接続ストレージ(NAS)(204a)および(204b)を含むネットワーク(200)を示す略図である。例示的な実施形態において、システム(202a)、(202b)、および(202c)は、データ記憶を管理し、ネットワーク接続ストレージ(NAS)(204a)および(204b)に記憶されたデータに対するデータアクセスを最適化することができる。数学モデルがこのデータに対して使用され、コンピュータシステム(202a)および(202b)、ならびに携帯電話およびパーソナルデータアシスタントシステム(202c)にわたって分散並列処理を使用して評価され得る。コンピュータシステム(202a)および(202b)、ならびに携帯電話およびパーソナルデータアシスタントシステム(202c)はさらに、ネットワーク接続ストレージ(NAS)(204a)および(204b)に記憶されたデータの適応データ再構築に対して並列処理を提供することができる。図5は一例のみを例示しており、多種多様な他のコンピュータのアーキテクチャおよびシステムは、本開示の様々な実施形態と連動して使用され得る。例えば、ブレードサーバーは並列処理を提供するために使用することができる。プロセッサブレードは、並列処理を提供するためにバックプレーンを介して接続可能である。ストレージも、バック

【0503】

いくつかの例示的な実施形態では、プロセッサは、別のメモリ空間を維持することができる。他のプロセッサによる並列処理のために、ネットワークインターフェース、バックプレーン、または他のコネクタを介してデータを送信することができる。他の実施形態では、プロセッサのいくつかまたはすべては、共有の仮想アドレスメモリ空間を使用することができる。

【0504】

図6は、例示的な実施形態にかかる共有の仮想アドレスメモリ空間を使用する、マルチ

10

20

30

40

50

プロセッサコンピュータシステム（300）のブロック図である。システムは、共有メモリサブシステム（304）にアクセス可能な複数のプロセッサ（302a-f）を含む。システムは、メモリサブシステム（304）に複数のプログラマブルハードウェアのメモリアルゴリズムプロセッサ（MAP）（306a-f）を組み込む。各MAP（306a-f）は、メモリ（308a-f）および1つ以上のフィールドプログラマブルゲートアレイ（FPGA）（310a-f）を含み得る。MAPは設定可能な機能ユニットを提供し、特定のアルゴリズムまたはその一部は、各プロセッサと密接に協働して処理を行うためにFPGA（310a-f）に提供され得る。例えば、MAPは、データモデルに関する代数式を評価するために、および例示的な実施形態において適応データの再構成を実行するために使用され得る。この例では、各MAPは、このような目的のためにプロセッサすべてによって世界中からアクセス可能である。1つの構成において、各MAPは、関連するメモリ（308a-f）にアクセスするためにダイレクトメモリアクセス（DMA）を使用することができ、それにより、それぞれのマイクロプロセッサ（302a-f）とは無関係に、かつこれらから非同期的に、タスクを実行することが可能となる。この構成では、MAPは、アルゴリズムのパイプライン処理（*pipelining*）および並列実行のために別のMAPに結果を直接供給することができる。

10

【0505】

上記のコンピュータのアーキテクチャおよびシステムは、例に過ぎず、汎用のプロセッサ、コプロセッサ、FPGA、および他のプログラマブルロジックデバイス、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、ならびに他の処理素子および論理素子のあらゆる組み合わせを使用するシステムを含む、多種多様な他のコンピュータ、携帯電話、パーソナルデータアシスタントのアーキテクチャおよびシステムが、例示的な実施形態と連動して使用され得る。いくつかの実施形態では、コンピュータシステムのすべてまたは一部は、ソフトウェアまたはハードウェアにおいて実施され得る。ランダムアクセスメモリ、ハードドライブ、フラッシュメモリ、テープドライブ、ディスクアレイ、ネットワーク接続ストレージ（NAS）、他のローカルまたは分散データストレージデバイスおよびシステムを含む、あらゆる種類のデータストレージ媒体も例示的な実施形態と連動して使用され得る。

20

【0506】

例示的な実施形態では、コンピュータシステムは、上記または他のコンピュータのアーキテクチャおよびシステムのいずれかで実行するソフトウェアモジュールを使用して実施され得る。他の実施形態において、システムの機能は、ファームウェア、図11で言及されるようなフィールドプログラマブルゲートアレイ（FPGA）などのプログラム可能な論理回路、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、または他の処理素子および論理素子において、部分的または完全に実施され得る。例えば、セットプロセッサおよびオブティマイザは、図4に例示されるアクセラレータカード（122）などのハードウェアアクセラレータカードの使用を介してハードウェアアクセラレーションで実施され得る。

30

【0507】

定義

他に定義されない限り、本明細書で用いるすべての技術的および科学的な用語は、本開示の属する技術分野における当業者によって一般に理解されるものと同じ意味を有する。本明細書に記載されるものと同様または同等の方法および試薬が、開示された方法および組成物の実施に使用され得るが、ここでは例示的な方法および材料が記載されている。

40

【0508】

本明細書と添付の特許請求の範囲において使用されるように、単数形「a」、「an」、および「the」は、文脈が明確に別のことを指定していない限り、複数の指示対象を含む。したがって、例えば、「コンティグ」への言及は、複数のそのようなコンティグを含み、「染色体の物理的なレイアウトをプローブする」への言及は、当業者に既知の染色体およびその同等物の物理的なレイアウトをプローブする1つ以上の方法への言及を含む

50

。

【0509】

さらに、「および (and)」の使用は、特に明記しない限り、「および／または」を意味する。同様に、「含む (comprise)」、「含む (comprises)」、「含む (comprising)」、「含む (include)」、「含む (includes)」、および「含む (including)」は互換的なものであり、限定的なものとしては意図されていない。

【0510】

さらに、様々な実施形態の記載が「含む (comprising)」という用語を使用する場合、当業者は、いくつかの特定の例において、「～から実質的になる」または「～からなる」という用語を使用して実施形態が代替的に記載され得ることを理解するものと理解されたい。

10

【0511】

「配列決定リード」という用語は、本明細書で使用されるように、配列が判定された DNA の断片を指す。

【0512】

「コンティグ」という用語は、本明細書で使用されるように、DNA 配列の連続領域を指す。「コンティグ」は、当該技術分野で既知である任意の数の方法、例えば、重複配列について配列決定リードを比較すること、および／または、どの配列決定リードが連続する可能性が高いかを識別するために既知の配列のデータベースに対して配列決定リードを比較することにより判定され得る。

20

【0513】

本明細書で使用されるような「対象」という用語は、真核生物または原核生物を指すことができる。

【0514】

本明細書で使用されるような「リード対 (read pair) または (read-pair)」という用語は、配列情報を提供するために結合される2つ以上の要素を指すことができる。場合によっては、リード対の数は、マッピング可能なリード対の数を指す場合がある。他の場合において、リード対の数は、生成されたリード対の総数を指す場合がある。

30

【0515】

本明細書で使用される「安定化された」という用語は、保存されているか、または分解から保護されている試料を記載することができる。場合によっては、安定化された試料は、固定剤またはクロスリンク剤でクロスリンクまたは処理される。場合によっては、安定化された試料は、ホルムアルデヒド、ホルマリン、パラホルムアルデヒド、グルタルアルデヒド、四酸化オスミウムなどで処理される。

【0516】

本明細書で使用される「約」という用語は、別段の指定がない限り、その数のプラスまたはマイナス10%の数を含む値の範囲として、その数を記載することができる。

【0517】

本明細書で使用される場合、「核酸の露出した内部末端」は、フェーズまたは局所的な三次元構造情報を容易にするために末端隣接核酸配列情報にアクセスするために導入されたものなど、安定化または非安定化された核酸に導入された切断部位の生成によって生成された露出した末端を指すことができる。

40

【0518】

本明細書で使用される場合、用語「約」数字は、その数の+/-10%に及ぶ範囲を指し、「約」範囲は、規定された範囲限界よりも10%低い範囲～規定された範囲限界よりも10%大きい範囲を指す。

【0519】

本明細書で使用される場合、リンカー上の配列セグメントは、その配列の同定が、第2

50

のパーティションまたは起源細胞を除外して、特定の第1のパーティションまたは起源細胞に、隣接する核酸配列を割り当てることを容易にするとき、パーティション指定または細胞指定である。識別配列は、場合によっては、パーティションまたは細胞に固有であり、したがって、他のすべての細胞と区別し、これが技術的に実行可能である場合、固有のタグは下流の分析を容易にする。しかしながら、固有の配列はすべての場合において必要とされるわけではない。場合によっては、冗長なバーコード化は、計算的に下流で解決され、それにもかかわらず、固有でないタグは、第1のパーティションまたは細胞の核酸を、第2のパーティションまたは細胞から区別するのに充分である。

【0520】

本明細書で使用される場合、クラスターは、複数の別個の末端隣接配列または配列タグがマッピングされる核酸参照の領域である。場合によっては、第2の領域に対する1つの領域の近接性は、少なくとも部分的に、第2のクラスターのクラスター成分とペアエンドのリードにおいて同時に生じる第1のクラスターのクラスター成分の数を計数することによって評価される。

【実施例】

【0521】

以下の実施例は、本発明の様々な実施形態を例示する目的で与えられ、いかなる方法でも本発明を制限するようには意図されていない。本明細書に記載される方法とともに、本実施例は、好ましい実施形態を代表するものであり、例示的なものであり、および、本発明の範囲を限定することを意図したものではない。請求項の範囲によって定義される本発明の精神内に包含されるその変化および他の使用は、当業者に想定される。

【0522】

実施例1：試料調製

試料の調製には、試料の種類：細胞または組織に応じて2つの別々のプロトコルがある。溶解物定量化工程は、両方の試料の種類について同じである。試料の調製には2時間かけるべきである。

【0523】

注記：10倍のHiC洗浄緩衝液、10倍の脱クロスリンク緩衝液（Crosslink Reversal Buffer）、および20%SDSは、保存時に沈殿した可能性がある。沈殿物がもはや見えなくなるまで、溶液を37℃で15分間インキュベートする。使用前にボルテックスして混合する。10倍のHiC洗浄緩衝液を超純水で1倍に希釈する。室温で保存する。試料あたり約15mLの1倍のHiC洗浄緩衝液が必要である。1倍のHiC洗浄緩衝液もプロトコルの残りを通して使用することができる。1倍のHiC洗浄緩衝液は室温で2ヶ月間安定している。10倍の脱クロスリンク緩衝液を超純水で1倍に希釈する。室温で保存する。試料あたり約1mLの1倍の脱クロスリンク緩衝液が必要である。1倍の脱クロスリンク緩衝液も近接ライゲーションプロトコルに使用することができる。1倍の脱クロスリンク緩衝液は、室温で2ヶ月間安定している。1.5mLチューブについて攪拌熱ミキサーを1250rpmに設定しなければならない。使用前に氷上で緩衝液を解凍し、ボルテックスすることを含む、優良試験所基準を使用する。

【0524】

細胞のためのプロトコル

注記：10×10⁶細胞を出発物質として使用して、洗浄中の損失を考慮することが推奨される。10×10⁶未満の細胞が利用可能である場合、低入力プロトコルを参照されたい。開始前に、新鮮な1倍のヌクレアーゼ消化緩衝液を調製し、室温で保存する。1倍のヌクレアーゼ消化緩衝液は、室温で1日安定している。1倍のヌクレアーゼ消化緩衝液を調製するために、140μLの超純水、20μLの10倍のヌクレアーゼ消化緩衝液、20μLの100mMのMnCl₂、20μLの10%トリトンを混合する。

【0525】

細胞を回収し、1倍のPBSで洗浄する。細胞を計数し、10×10⁶細胞を等分し、2000×gで5分間遠心分離する。上清を注意深く除去する。ペレットを5mLの1倍

のPBSおよび135 μ lの37%ホルムアルデヒドに再懸濁する。試料を5mlチューブに移し、細胞が沈降しないような速度にて室温で10分間回転させる。チューブを2000 $\times$ gで5分間遠心分離する。細胞ペレットが緩む可能性があるため、注意しながら上清を注意深く除去する。ペレットをHiC洗浄緩衝液で洗浄し、最初に200 μ lで凝集塊を破壊し、次いで残りの4.8mlを添加し、上下にピペティングしてペレットを完全に再懸濁する。チューブを2000 $\times$ gで5分間回転させ、上清を注意深く除去する。合計2回の洗浄のために洗浄工程を繰り返す。2回目の洗浄液を除去した後、ペレットを1倍のHiC洗浄緩衝液1mlに再懸濁し、ペレットを再懸濁する。細胞を計数し、 1×10^6 細胞を3つの別々のチューブに加え、残りの細胞を -80°C で凍結したペレットとして保存する。3つのチューブを2000 $\times$ gで5分間遠心分離し、上清を除去する。各チューブ中のペレットを50 μ lの1倍のヌクレアーゼ消化緩衝液（新しく調製した）に再懸濁する。チューブを、1250 rpmの攪拌熱ミキサーにおいて2分間 30°C に予熱する。7.5 μ lのヌクレアーゼ酵素混合物を含む新鮮な1.5mlチューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で2分間予熱する。予熱したヌクレアーゼ酵素混合物を、以下のように予熱した各チューブに、0.5 μ lを第1のチューブに移し、2.0 μ lを第2のチューブに移し、4.0 μ lを第3のチューブに移す。チューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で正確に30分間インキュベートする。5 μ lの0.5MのEDTAを添加して混合することによって、ヌクレアーゼ反応を停止させる。3 μ lの20%SDSを添加して細胞を溶解し、細胞を1250 rpmで、攪拌熱ミキサー中 30°C で5分間インキュベートする。

10

20

【0526】

組織のためのプロトコル

注記：60mgの組織を使用することが推奨される。60mg未満の組織の細胞が利用可能である場合、低入力プロトコルを参照されたい。開始前に、新鮮な1倍のヌクレアーゼ消化緩衝液を調製し、室温で保存する。1倍のヌクレアーゼ消化緩衝液は、室温で1日安定している。1倍のヌクレアーゼ消化緩衝液を調製するために、140 μ lの超純水、20 μ lの10倍のヌクレアーゼ消化緩衝液、20 μ lの100mMのMnCl₂、20 μ lの10%トリトンを混合する。

【0527】

少なくとも60mgの凍結組織を秤量し、図1Aおよび図1Bで例示される稠度まで液体窒素中で乳鉢および乳棒を用いて微粉末に粉碎し、図1Aは不十分な組織粉碎を示し、図1Bは十分な組織粉碎を示す。破壊した組織を、5mlの1倍のPBSおよび135 μ lの37%ホルムアルデヒドを含む5mlチューブに移す。チューブを室温で10分間回転させる。チューブを2000 $\times$ gで5分間遠心分離し、上清を注意深く除去する。組織がペレット化しない場合、チューブを最大速度で5分間回転させる。ペレットを200 μ lの洗浄緩衝液に再懸濁し、次いで4.8mlの1倍のHiC洗浄緩衝液を加える。チューブを2000 $\times$ gで5分間遠心分離し、上清を除去する。洗浄工程を2回行い、最終ペレットを1倍のHiC洗浄緩衝液1mlに再懸濁する。再懸濁した細胞を200 μ mフィルターに通して新鮮な5mlチューブに入れ、必要に応じてフィルターを交換する。さらに2mlの1倍のHiC洗浄緩衝液を200 μ mフィルターに通す。試料を3つの別々のチューブ中で3つの1mlアリコートに分離し、各アリコートは20mgの組織に対応する。余剰組織をペレット化し、 -80°C で保存することができる。3つのチューブを2000 $\times$ gで5分間遠心分離し、上清を除去する。各チューブ中のペレットを50 μ lの1倍のヌクレアーゼ消化緩衝液（新しく調製した）に再懸濁する。チューブを、1250 rpmの攪拌熱ミキサーにおいて2分間 30°C で予熱する。7.5 μ lのヌクレアーゼ酵素混合物を含む新鮮な1.5mlチューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で2分間予熱する。予熱したヌクレアーゼ酵素混合物を、以下のように予熱した各チューブに、0.5 μ lを第1のチューブに移し、2.0 μ lを第2のチューブに移し、4.0 μ lを第3のチューブに移す。チューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で正確に30分間インキュベートする。5 μ lの0.5MのEDTAを添加して

30

40

50

混合することによって、ヌクレアーゼ反応を停止させる。3 μ l の 20 % SDS を添加して細胞を溶解し、細胞を 1250 rpm で、攪拌熱ミキサー中 30 °C で 5 分間インキュベートする。

【0528】

実施例 2：溶解物の定量化

注記：溶解物の定量化には 2 時間かけなければならない。80 % エタノールを、SPRI Select ビーズを用いた DNA 精製のために新たに調製する。定量化工程は、2 つの目的：近接ライゲーションステップで使用する試料の体積を決定すること、試料調製で得た 3 つのチューブのどれを近接ライゲーションで使用するかを決定することを有する。

【0529】

各チューブからの 2 μ l の溶解物を 18 μ l の 1 倍の HiC 洗浄緩衝液と混合することによって、各溶解物を 1 : 10 に希釈する。希釈されていない溶解物を -80 °C で保存する。2.5 μ l の各溶解物を、50 μ l の脱クロスリンク緩衝液および 1.5 μ l のプロテイナーゼ K を含むチューブに移す。混合物をピペティングで混合し、1250 rpm で攪拌熱ミキサーで 55 °C で 15 分間、次いで 68 °C で 45 分間インキュベートする。100 μ l の SPRI Select ビーズを各チューブに加え、ボルテックスして再懸濁させ、スピンドウンさせ、磁石から離して室温で 5 分間インキュベートする。チューブを磁石中に 5 分間、または溶液が透明に見え、ビーズが完全に分離されるまで置く。上清を除去し、ビーズを、磁石からチューブを取り外すことなく、80 % エタノールで 1 分間 2 回洗浄する。2 回目の洗浄後、チューブをスピンドウンさせ、磁石上に 1 分間置いた。残りのエタノールをピペットで除去する。ビーズを、エタノールが残らなくなるまで磁石上で 5 分間空気乾燥させるが、過剰乾燥させることはない。チューブを磁石から取り出し、ビーズを 10 μ l の TE 緩衝液 (pH 8.0) に再懸濁する。チューブをボルテックスし、スピンドウンさせ、磁石上に 1 分間置く。8 μ l の上清を新鮮なチューブに移す。Qubit Fluorometer および Qubit dsDNA HS キットを用いて試料を定量化する。濃度はスプレッドシートに記録される。断片サイズ分布は、TapeStation D5000 または D5000 HS ScreenTape を用いて決定される。D5000 HS ScreenTape を使用する場合、試料を 1 ng / μ l に希釈しなければならない。領域：領域 1 100 ~ 2500 bp、領域 2 100 ~ 600 bp、および領域 3 600 ~ 2500 bp を TapeStation System で分析する。全パーセントを計算し、記録する。データは図 2 に示すように現れ、合計 76.33 % が 100 bp ~ 2500 bp であり、合計 28.82 % が 100 bp ~ 600 bp であり、合計 47.82 % が 600 bp ~ 2500 bp である。

【0530】

1000 ng に対応する試料の体積を計算する。クロマチン消化効率 (CDE) およびクロマチン消化指数 (CI) を計算する。どの試料が QC メトリックに合格するかを決定する。

【0531】

実施例 3：近接ライゲーション

注記：近接ライゲーションには 5.5 時間をかけなければならない。攪拌熱ミキサーを 1.5 ml チューブについて 1250 rpm に設定する。試料を磁石の上に置く際、時間を経過させて、上清を除去する前に溶液を完全に透明にする。新鮮な架橋ライゲーション混合物 (50 μ l) を調製し、使用前に氷上で保存する。50 μ l の架橋ライゲーション混合物を調製するために、10 μ l の 5 $\times$ ライゲーション緩衝液、5 μ l の架橋、および 35 μ l の超純水を混合する。80 % エタノールを、SPRI Select ビーズを用いた DNA 精製のために新たに調製する。

【0532】

クロマチン捕捉ビーズへのクロマチンの結合

クロマチン捕捉ビーズを室温にして、使用前にボルテックスする。100 μ l のクロマチン捕捉ビーズを、計算された実施例 1 からの試料 1000 ng を含む 1.5 ml チュー

10

20

30

40

50

ブに移す。試料をピペッティングによって混合し、磁石から離して室温でインキュベートする。チューブを磁石上に5分間、または溶液が透明になりビーズが分離するまで置く。上清を除去する。チューブを磁気ラックから取り出し、ビーズを150 μ lの1倍のHi-C洗浄緩衝液で洗浄し、10回のピペッティングによって混合し、その後、1分間磁石上に置いた後、上清を除去する。洗浄工程を1回繰り返す。

【0533】

末端平滑化

チューブを磁気ラックから取り出し、50 μ lの末端平滑化緩衝液および3.5 μ lの末端平滑化酵素混合物を添加する。試料をピペッティングによって混合し、22℃で30分間、その後、65℃で30分間、1250 rpmの攪拌熱ミキサーにおいてインキュベートする。チューブを室温に到達させ、次いで磁気ラック上に1分間、または溶液が透明に見えてビーズが分離するまで置く。上清を除去する。チューブを磁気ラックから取り出し、ビーズを150 μ lの1倍のHi-C洗浄緩衝液で1回洗浄し、ピペッティングによって混合する。次いで、チューブを磁気ラック上に1分間再度置き、上清を除去する。

【0534】

架橋ライゲーション

チューブを磁気ラックから取り出し、50 μ lの架橋ライゲーション混合物を1 μ lのT4 DNAリガーゼと共に添加する（新たに作製する）。試料をピペッティングによって混合し、22℃で30分間インキュベートする。チューブを磁気ラック上に1分間、または溶液が透明になるまで置く。上清を除去し、チューブを磁気ラックから取り出す。ビーズを150 μ lの1倍のHi-C洗浄緩衝液で再懸濁し、ピペッティングによって混合する。チューブを磁石上に1分間置き、上清を除去する。

【0535】

凝集体内ライゲーション

チューブを磁気ラックから取り出し、50 μ lの凝集体内ライゲーション緩衝液および2 μ lの凝集体内ライゲーション酵素混合物をビーズに加える。試料をピペッティングによって混合し、攪拌熱ミキサーで22℃で1時間インキュベートする。チューブを磁気ラック上に1分間、または溶液が透明に見え、ビーズが分離するまで置く。上清を除去する。

【0536】

脱クロスリンク (Crosslink Reversal)

チューブを磁気ラックから取り出し、50 μ lに加え1.5 μ lのプロテイナーゼKをビーズに加える。試料をピペッティングで混合し、1250 rpmで攪拌熱ミキサーで55℃で15分間、次いで68℃で45分間インキュベートする。

【0537】

SPRIselectビーズ上でのDNA精製

SPRIselectビーズを30秒間ボルテックスして再懸濁させる。35 μ lの再懸濁ビーズを1.5 mlの試料チューブに加える。試料をボルテックスし、スピンドウンさせ、磁石から離して室温で5分間インキュベートする。チューブを磁石上に5分間、または溶液が透明に見え、ビーズが分離するまで置く。上清を除去する。チューブを磁石上に残して、150 μ lの80%エタノールで2回洗浄する。ビーズをこれらの洗浄液に再懸濁せず、エタノールを添加し、1分間インキュベートし、エタノールを除去する。2回目の洗浄後、チューブをスピンドウンさせ、磁石上に1分間置き、ピペットを使用してエタノールの最後を除去する。ビーズを、エタノールが残らなくなるまで磁石上で5分間空気乾燥させるが、ビーズは乾燥し過ぎない。試料を磁石から取り出し、52 μ lのTE緩衝液 (pH 8.0) を加える。試料をボルテックスし、スピンドウンさせ、磁石から離して室温で5分間インキュベートする。チューブをスピンドウンさせ、磁石上に1分間置く。50 μ lの上清を新鮮な1.5 mlチューブに移す。ビーズを廃棄する。Qubit FluorometerおよびQubit dsDNA HSキットを用いて試料を定量化する。ライブラリー調製工程に進むには200 ngが必要である。精製DNAを20℃

10

20

30

40

50

で最大6ヶ月保存する。

【0538】

実施例4：ライブラリー調製

注記：ライブラリー調製プロトコルはフラグメンテーションを必要とせず、約2時間かけなければならない。

【0539】

末端修復

注記：末端修復緩衝液は、保存中に沈殿することがあるが、目に見える沈殿物がなくなるまで37℃で少なくとも10分間インキュベートすべきである。250mMのDTTをピペティングによって混合して、使用前に完全に混合する。

【0540】

0.2mLのPCRチューブを、48μLの精製試料、7μLの末端修復緩衝液、3μLの末端修復酵素混合物、および0.5μLの250mMのDTTで調製する。混合物をピペティングによって混合し、スピンドウンさせる。試料をサーマルサイクラー中で20℃で30分間、その後、65℃で30分間インキュベートする。試料を12℃に保持する。

【0541】

アダプターライゲーションおよびUSER消化

試料を含有する0.2mLのPCRチューブを、Illumina用の2.5μLのアダプター、1μLのライゲーションエンハンサー、および30μLのライゲーション酵素混合物と混合する。試料をピペティングによって混合し、チューブをスピンドウンさせる。試料をサーマルサイクラー中20℃で15分間インキュベートし、12℃で保持する。3μLのUSER酵素混合物をPCRチューブに加える。試料をピペティングによって混合し、スピンドウンさせる。試料をサーマルサイクラー中37℃で15分間インキュベートし、12℃で保持する。

【0542】

DNA精製

SPRIselectビーズを30秒間ボルテックスして再懸濁させる。80μLの再懸濁ビーズをPCRチューブに加える。試料をボルテックスし、スピンドウンさせ、磁石から離して室温で5分間インキュベートする。チューブを磁石上に5分間、または溶液が透明に見え、ビーズが分離するまで置く。上清を除去する。チューブを磁石上に残して、150μLの80%エタノールで2回洗浄する。ビーズをこれらの洗浄液に再懸濁せず、エタノールを添加し、1分間インキュベートし、エタノールを除去する。2回目の洗浄後、チューブをスピンドウンさせ、磁石上に1分間置き、ピペットを使用してエタノールの最後を除去する。ビーズを、エタノールが残らなくなるまで磁石上で5分間空気乾燥させるが、ビーズは乾燥し過ぎない。試料を磁石から取り出し、100μLのTE緩衝液(pH8.0)を加える。試料をボルテックスし、スピンドウンさせ、磁石から離して室温で5分間インキュベートする。チューブをスピンドウンさせ、磁石上に1分間置く。95μLの上清を新鮮な1.5mLチューブに移す。ビーズを廃棄する。精製DNAを20℃で一晩保存する。

【0543】

実施例5：ライゲーション捕捉および増幅

注記：ライゲーション捕捉および増幅プロトコルには2時間かけなければならない。

【0544】

ストレプトアビジンビーズ調製

注記：この工程はいかなるDNA試料も含まない。

【0545】

ストレプトアビジンビーズをボルテックスして再懸濁させる。25mLの再懸濁ストレプトアビジンビーズを1.5mLチューブに移す。ストレプトアビジンビーズを含むチューブを磁石上に5分間置き、上清を除去する。チューブを磁気ラックから取り出し、スト

10

20

30

40

50

レプトアビジンビーズを200 μ lのTWB (Red Label) で洗浄し、ピペッティングによって混合する。試料を磁気ラック上に1分間置き、上清を除去する。洗浄工程を1回繰り返す。次いで、ストレプトアビジンビーズを100 μ lの2 $\times$ NTP (Yellow Label) に再懸濁し、ピペッティングによって混合する。

【0546】

ライゲーション捕捉

95 μ lの精製DNAを、100 μ lの2 $\times$ NTPに再懸濁したストレプトアビジンビーズと共に1.5mlチューブに移す。チューブを10秒間ボルテックスし、スピンドウンさせる。混合物を攪拌熱ミキサーにおいて25℃で30分間インキュベートする。

【0547】

ストレプトアビジンビーズ上での試料の洗浄

注記：各洗浄について、チューブを磁気ラックから取り出し、指示された緩衝液をビーズに加え、ビーズを再懸濁し、チューブを磁石上に1分間置き、各洗浄の間にすべての上清を除去するように注意しながら、上清を除去する。

【0548】

チューブをスピンドウンさせ、次いで磁石上に1分間置き、上清を除去する。ビーズを200 μ lのLWBで1回洗浄する。ビーズを200 μ lのNWBで2回洗浄する。ビーズを200 μ lの1倍のHiC洗浄緩衝液で2回洗浄する。

【0549】

インデックスPCR

注記：すべてのPCR酵素およびマスターミックスがストレプトアビジンビーズの存在下での増幅に適合性があるわけではないため、供給されたPCR Ready Mixを使用する。

【0550】

最後の洗浄液を除去した後、チューブを磁気ラックから取り出し、25 μ lのHot Start PCR Ready Mix、5 μ lのユニバーサルPCRプライマー (Universal PCR Primer)、5 μ lのインデックスプライマー (Index Primer) (各試料に固有)、ならびに15 μ lのDNaseおよびRNaseフリーの蒸留水をビーズに加える。混合物をピペッティングによって混合し、0.2mlのPCRチューブに移す。このチューブをスピンドウンさせ、サーモサイ클ラーに入れて、以下のとおりのプログラム：98℃で3分、12サイクルの(98℃で20秒、65℃で30秒、72℃で30秒)、72℃で1分、12℃で保持を実行する。

【0551】

サイズ選択

PCRチューブをスピンドウンし、磁石上に1分間置く。47 μ lの上清を1.5mlチューブに移し、ビーズを廃棄する。53 μ lのTE緩衝液 (pH 8.0) をチューブに添加して、総体積を100 μ lにする。SPRIselectビーズを30秒間ボルテックスし、45 μ lの再懸濁したSPRIselectビーズを、試料を含む1.5mlのチューブに加える。混合物をボルテックスして再懸濁させ、スピンドウンさせ、磁石から離して室温で10分間インキュベートする。チューブをスピンドウンし、磁石上に5分間置く。145 μ lの上清を新しい1.5mlチューブに移し、ビーズを廃棄する。35 μ lのSPRIselectビーズを1.5mlチューブに加え、ボルテックスして再懸濁させ、スピンドウンさせ、磁石から離して室温で10分間インキュベートする。チューブをスピンドウンし、磁石上に5分間置く。上清を除去する。チューブを磁石上に残し、ビーズを200 μ lの80%エタノールで2回洗浄する。ビーズは、これらの洗浄のために再懸濁されない。チューブをスピンドウンし、磁石上に1分間置く。10 μ lのピペットチップを使用して、微量のエタノールを除去する。ビーズを、残留エタノールが残らなくなるまで5分間磁石上で空気乾燥させるが、過剰乾燥させない。ビーズを30 μ lのTE緩衝液 (pH 8.0) に再懸濁し、ピペッティングにより混合する。チューブをスピンドウンし、磁石から離して室温で2分間インキュベートする。チューブをスピンドウンし、

10

20

30

40

50

磁石上に1分間置く。28 μ lの上清を新しい1.5mlチューブに移す。このチューブはライブラリーを含む。サイズ選択されたライブラリーを、Qubit FluorometerおよびQubit dsDNA HSキットを用いて定量化する。少なくとも60 ngのDNAが回収される。60 ng未満のDNAが回収される場合、ライブラリーを廃棄する。TapeStationまたはBioanalyzerを使用して、サイズ選択されたライブラリーのサイズ分布を検証し、ライブラリーのサイズ範囲は350 bp ~ 1000 bpである。ライブラリーを-20℃で最大6ヶ月保存する。

【0552】

実施例6：低入力サンプル調製

【0553】

これは、推奨入力を利用可能でないときに使用される。低試料入力により、複雑性の低い最終ライブラリーがしばしば得られる。

【0554】

細胞

利用可能な細胞の数を使用し、細胞に対する実施例1の方法をヌクレアーゼ工程まで使用する。ヌクレアーゼ工程において、0.1 μ lの予熱したヌクレアーゼ酵素混合物を第1のチューブに加え、0.5 μ lの予熱したヌクレアーゼ酵素混合物を第2のチューブに加え、2.0 μ lのヌクレアーゼ酵素混合物を第3のチューブに加える。

【0555】

組織

利用可能な組織の量（少なくとも5mg）を使用し、組織に対する実施例1の方法をヌクレアーゼ工程まで使用する。ヌクレアーゼ工程において、0.1 μ lの予熱したヌクレアーゼ酵素混合物を第1のチューブに加え、0.5 μ lの予熱したヌクレアーゼ酵素混合物を第2のチューブに加え、2.0 μ lのヌクレアーゼ酵素混合物を第3のチューブに加える。

【0556】

実施例7：インデックスプライマー

以下のインデックスプライマーを使用する。

【0557】

【表1】

表1: インデックスプライマー	
インデックスプライマー	配列
インデックスプライマー2	CGATGT
インデックスプライマー4	TGACCA
インデックスプライマー5	ACAGTG
インデックスプライマー6	GCCAAT
インデックスプライマー7	CAGATC
インデックスプライマー8	ACTTGA
インデックスプライマー12	CTTGTA
インデックスプライマー19	GTGAAA

【0558】

インデックスプライマーは以下のスキームに従って選択される。

【0559】

【表2】

表 2: インデックスプライマー選択	
ライブラリーの数	インデックスプライマーの組み合わせ
2	6 及び 12
2	5 及び 19
3	2, 7, 及び 19
3	2-プレックスオプションのいずれかと任意の他のインデックスプライマー
4	5, 6, 12, 及び 19
4	3-プレックスオプションのいずれかと任意の他のインデックスプライマー

10

20

【0560】

実施例 8：試料調製－MNase

試料の調製には、試料の種類：細胞または組織に応じて2つの別々のプロトコルがある。溶解物の定量化工程は、両方の試料の種類について同じである。試料の調製には2時間かけるべきである。

【0561】

注記：10倍のHiC洗浄緩衝液、10倍の脱クロスリンク緩衝液、および20%SDSは、保存時に沈殿した可能性がある。沈殿物がもはや見えなくなるまで、溶液を37℃で15分間インキュベートする。使用前にボルテックスして混合する。10倍のHiC洗浄緩衝液を超純水で1倍に希釈する。室温で保存する。試料あたり約15mLの1倍のHiC洗浄緩衝液が必要である。1倍のHiC洗浄緩衝液もプロトコルの残りを通して使用することができる。1倍のHiC洗浄緩衝液は室温で2ヶ月間安定している。10倍の脱クロスリンク緩衝液を超純水で1倍に希釈する。室温で保存する。試料あたり約1mLの1倍の脱クロスリンク緩衝液が必要である。1倍の脱クロスリンク緩衝液も近接ライゲーションプロトコルに使用することができる。1倍の脱クロスリンク緩衝液は、室温で2ヶ月間安定している。1.5mLチューブについて攪拌熱ミキサーを1250rpmに設定しなければならない。使用前に氷上で緩衝液を解凍し、ボルテックスすることを含む、優良試験所基準を使用する。

30

40

【0562】

細胞のためのプロトコル

10 x 10⁶細胞を出発物質として使用して、洗浄中の損失を考慮することが推奨される。10 x 10⁶未満の細胞が利用可能である場合、低入力プロトコルを参照されたい。開始前に、新鮮な1倍のMNase消化緩衝液を調製し、室温で保存する。1倍のMNase消化緩衝液は、室温で1日安定している。1倍のMNase消化緩衝液を調製するために、140μLの超純水、20μLの10倍のMNase消化緩衝液、20μLの100mMのMnCl₂、20μLの10%トリトンを混合する。

50

【0563】

細胞を回収し、1倍のPBSで洗浄する。細胞を計数し、 10×10^6 細胞を等分し、 $2000 \times g$ で5分間遠心分離する。上清を注意深く除去する。ペレットを5mlの1倍のPBSおよび135 μ lの37%ホルムアルデヒドに再懸濁する。試料を5mlチューブに移し、細胞が沈降しないような速度で室温で10分間回転させる。チューブを $2000 \times g$ で5分間遠心分離する。細胞ペレットが緩む可能性があるため、注意しながら上清を注意深く除去する。ペレットをHiC洗浄緩衝液で洗浄し、最初に200 μ lで凝集塊を破壊し、次いで残りの4.8mlを添加し、上下にピペティングしてペレットを完全に再懸濁する。チューブを $2000 \times g$ で5分間回転させ、上清を注意深く除去する。合計2回の洗浄のために洗浄工程を繰り返す。2回目の洗浄液を除去した後、ペレットを1倍のHiC洗浄緩衝液1mlに再懸濁し、ペレットを再懸濁する。細胞を計数し、 1×10^6 細胞を3つの別々のチューブに加え、残りの細胞を -80°C で凍結したペレットとして保存する。3つのチューブを $2000 \times g$ で5分間遠心分離し、上清を除去する。各チューブ中のペレットを50 μ lの1倍のMNase消化緩衝液（新しく調製した）に再懸濁する。チューブを、1250 rpmの攪拌熱ミキサーにおいて2分間 30°C に予熱する。7.5 μ lのMNase酵素混合物を含む新鮮な1.5mlチューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で2分間予熱する。予熱したMNase酵素混合物を、以下のように予熱した各チューブに、0.5 μ lを第1のチューブに移し、2.0 μ lを第2のチューブに移し、4.0 μ lを第3のチューブに移す。チューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で正確に30分間インキュベートする。5 μ lの0.5 MのEDTAを添加して混合することによって、MNase反応を停止させる。3 μ lの20%SDSを添加して細胞を溶解し、細胞を1250 rpmで攪拌熱ミキサー中 30°C で5分間インキュベートする。

10

20

【0564】

組織のためのプロトコル

注記：60mgの組織を使用することが推奨される。60mg未満の組織の細胞が利用可能である場合、低入力プロトコルを参照されたい。開始前に、新鮮な1倍のMNase消化緩衝液を調製し、室温で保存する。1倍のMNase消化緩衝液は、室温で1日安定している。1倍のMNase消化緩衝液を調製するために、140 μ lの超純水、20 μ lの10倍のMNase消化緩衝液、20 μ lの100mMのMnCl₂、20 μ lの10%トリトンを混合する。少なくとも60mgの凍結組織を秤量し、図1Aおよび図1Bで例示される稠度まで液体窒素中で乳鉢および乳棒を用いて微粉末に粉碎し、図1Aは不十分な組織粉碎を示し、図1Bは十分な組織粉碎を示す。破壊した組織を、5mlの1倍のPBSおよび135 μ lの37%ホルムアルデヒドを含む5mlチューブに移す。チューブを室温で10分間回転させる。チューブを $2000 \times g$ で5分間遠心分離し、上清を注意深く除去する。組織がペレット化しない場合、チューブを最大速度で5分間回転させる。ペレットを200 μ lの洗浄緩衝液に再懸濁し、次いで4.8mlの1倍のHiC洗浄緩衝液を加える。チューブを $2000 \times g$ で5分間遠心分離し、上清を除去する。洗浄工程を2回行い、最終ペレットを1倍のHiC洗浄緩衝液1mlに再懸濁する。再懸濁した細胞を200 μ mフィルターに通して新鮮な5mlチューブに入れ、必要に応じてフィルターを交換する。さらに2mlの1倍のHiC洗浄緩衝液を200 μ mフィルターに通す。試料を3つの別々のチューブ中で3つの1mlアリコートに分離し、各アリコートは20mgの組織に対応する。余剰組織をペレット化し、 -80°C で保存することができる。3つのチューブを $2000 \times g$ で5分間遠心分離し、上清を除去する。各チューブ中のペレットを50 μ lの1倍のMNase消化緩衝液（新しく調製した）に再懸濁する。チューブを、1250 rpmの攪拌熱ミキサーにおいて2分間 30°C に予熱する。7.5 μ lのMNase酵素混合物を含む新鮮な1.5mlチューブを、1250 rpmの攪拌熱ミキサーにおいて 30°C で2分間予熱する。予熱したMNase酵素混合物を、以下のように予熱した各チューブに、0.5 μ lを第1のチューブに移し、2.0 μ lを第2のチューブに移し、4.0 μ lを第3のチューブに移す。チューブを、1250 rpmの攪拌

30

40

50

熱ミキサーにおいて30℃で正確に30分間インキュベートする。5μlの0.5MのEDTAを添加して混合することによって、MNase反応を停止させる。3μlの20%SDSを添加して細胞を溶解し、細胞を1250rpmで攪拌熱ミキサー中30℃で5分間インキュベートする。

【0565】

実施例9：MNase-Cライブラリーからの結果

本明細書中の方法を用いて近接ライゲーションライブラリーを調製し、配列決定してロングレンジ情報の尺度を決定した。図7は、DNase調製ライブラリー（「DNase-C」）、ならびに異なるクロスリンキング剤と比較したMNase調製ライブラリー（「MNase-C」）を使用したリード分離を示す。各バーについて、最下部のセグメントは、ゲノムにおいて1kb超離れたリード対のパーセンテージを示し、中間のセグメントは、ゲノムにおいて1kb未満離れたリード対のパーセンテージを示し、最上部のセグメントは、染色体間相互作用を有するリード対のパーセンテージを示す。左から右へ、バーは、1）ホルムアルデヒドクロスリンカーを含むDNase、2）DSGクロスリンカーを含むDNase、3）10分間反応させたEGSクロスリンカーを含むMNase、4）30分間反応させたEGSクロスリンカーを含むMNase、5）10分間反応させたDSGクロスリンカーを含むMNaseの結果を示す。図8は、DNase-C調製ライブラリー（DSGおよびホルムアルデヒドクロスリンカーの両方を含む）と比較した、MNase-C調製ライブラリー（DSGおよびEGSクロスリンカーの両方を含む）の1番染色体について計算した連鎖距離の分布を示す。DSGのより長いスパーサーアームを用いて調製されたDNase-Cライブラリーは、図7にも反映されるように、ホルムアルデヒドを用いて調製されたDNase-Cライブラリーと比較して、より大きい結合距離でリードのより大きい画分を示す。

【0566】

ゲノムワイドヌクレオソームマッピングの測定値も決定した。図9は、異なる量のMNaseおよび異なる消化時間：説明文に示されるように、0.05単位で30分間、0.5単位で30分間、2.5単位で20分間、および2.5単位で60分間で調製したライブラリーの高占有率CTCF結合部位周辺の相対的なリードカバレッジを示す。図10は、（左から右へ）0.05単位（U）で30分間（30'）、0.5単位（U）で30分間（30'）、2.5単位（U）で20分間（20'）、および2.5単位（U）で60分間（60'）において、調製されたライブラリーにおいて見られるモノヌクレオソーム：ジヌクレオソームの比を示し、一番右にラダーがある。モノヌクレオソーム：ジヌクレオソームの比は、（左から右へ）0.96、1.51、2.39、および4.86であった。

【0567】

実施例10：MNase消化試料によるHiC分析

使用者は生体試料中の核酸結合タンパク質の位置を高い精度で決定することを望む。生体試料は、核酸結合タンパク質を、それらが結合している核酸にクロスリンクするために、化学的固定剤を使用してクロスリンクされる。次いで、固定された試料を、タンパク質に結合していないすべての核酸を消化する小球菌ヌクレアーゼ（MNase）で消化する。次いで、MNase処理した核酸をDNAリガーゼで処理して、近接ライゲーション産物を得る。核酸を精製し、配列決定ライブラリーを作製する。配列決定ライブラリーを配列決定して、MNase消化試料のリード対を得る。核酸結合タンパク質に結合した核酸由来の配列のみが得られるので、リード対から、生体試料中の核酸結合タンパク質の局在化を高い精度で決定する。

【0568】

実施例11：MNase消化試料を含むHiChIP

MNase HiChIP分析を以下のように行った。細胞の試料をホルムアルデヒドおよびDSGクロスリンキングでクロスリンクし、MNaseでインサイチュで消化し、

R I P A 溶解緩衝液で溶解した。C C C T C 結合因子 (C T C F) および H 3 K 4 m e 3 修飾ヒストンに対する抗体を試料に接触させ、磁気ビーズを介してプルダウンした。末端平滑化、架橋ライゲーション、および凝集体内ライゲーションを含む近接ライゲーションを行った。次いで、クロスリンクングを脱クロスリンクさせ、D N A をクリーンアップし、配列決定ライブラリーを調製し、20~4000万の2×150bpリードで配列決定した。

【0569】

比較のため、C h I P - s e q 分析を、同じ抗体を有する試料の並行セットに対して行った。手順は同じであったが、クロスリンクング、消化、および近接ライゲーションの工程は行わなかった。

【0570】

図11は、University of California, Santa Cruz (UCSC) Genome BrowserからのEncyclopedia of DNA Elements (ENCODE) において報告されたピークと比較したC h I P - s e q および H i C h I P の結果を示す。陰性対照ライブラリーは、全ゲノムにわたってカバレッジを示すが、C T C F および H 3 K 4 m e 3 ライブラリーは、それらの結合部位についてのENCODEピークの位置と一致する遺伝子座における配列決定リードのパイルアップ (p i l e - u p s) を示す。これは、これらのライブラリーがC h I P - s e q シグナルを伝達し、リードがほぼ完全に、選択された抗体に対応するタンパク質が結合すると予想される領域からであることを示す。

【0571】

図12はH i C h I P 試料についてのC T C F 結合部位周辺の相対的なリードカバレッジを示す。相対的なリードカバレッジは約146bpの周期性を示し、これはヌクレオソーム中のヒストンの存在によってM N a s e 消化から保護されるD N A の領域と一致する。このことは、これらのライブラリーがM N a s e - C ライブラリーの予想される保護的なプロファイルを有することを示している。

【0572】

表3~5において、i n t e r . c h r は染色体間であるリード対のパーセンテージを表し (ペア中の各リードは異なる染色体にマッピングされる)、<1kbは1kb未満のゲノム上での分離距離にわたるリード対のパーセンテージを表し、>1kbは1kbを超えるゲノム上の分離距離にわたるリード対のパーセンテージを表し、f r a c M a p p e d は、マッピングされたリード対のパーセンテージを表し、p r e S e q A t 3 0 0 M は、3億リードの集団からの固有のリードの数を表す。

【0573】

表3は、M N a s e - C ライブラリーについての典型的なライブラリー品質管理 (Q C) メトリックを示す。染色体間リード対の割合は33%~35%であり、1kb未満の分離距離にわたるリード対の割合は7%~12%であり、1kbを超える分離距離にわたるリード対の割合は55%~58%であり、マッピングされた割合は67%~79%であり、3億のうちの固有のリード数は1.68億~2.38億の間である。表4に示すように、M N a s e H i C h I P ライブラリーのQ C メトリックは、M N a s e - C ライブラリーのQ C メトリックと比較的類似しており、染色体間リード対は21.6%~43.4%であり、1kb未満にわたるリード対は18.2%~31.2%であり、1kb超にわたるリード対は25.4%~56.2%であり、3億のうちの2.72億~2.79億の固有のリードである。比較すると、C h I P - s e q ライブラリーについてのQ C メトリックは、M N a s e - C ライブラリー (または他の近接ライゲーションライブラリー) のようではなく、染色体間リードは0.46%~1.65%であり、1kb未満に及ぶリード対は98.64%~99.45%であり、1kb超に及ぶリード対は0.09%~0.16%である。

【0574】

10

20

30

40

50

【表 3】

表 3: MNase-C の典型的なライブラリーQC メトリック					
試料 ID	inter.chr	<1kb	>1kb	fracMapped	preSeqAt300M
mdev3A	35%	7%	58%	67%	232,517,526
mdev3B	35%	8%	57%	68%	237,654,318
mdev3C	34%	11%	55%	77%	218,231,018
mdev3D	33%	12%	55%	77%	217,283,442
mdev3E	33%	11%	56%	78%	171,109,061
mdev3F	33%	11%	56%	79%	168,589,170

10

【 0 5 7 5 】

【表 4】

表 4: MNase HiChIP ライブラリーの QC メトリック				
	inter.chr	<1kb	>1kb	preSeqAt300M
陰性対照	21.6%	22.2%	56.2%	270,540,612
CTCF	40.3%	18.2%	41.5%	278,669,408
H3K36me3	43.4%	31.2%	25.4%	272,358,079

20

【 0 5 7 6 】

【表 5】

表 5: ChIP-seq ライブラリーQC メトリック				
	inter.chr	<1kb	>1kb	preSeqAt300M
陰性対照	1.23%	98.64%	0.13%	NA
CTCF	1.65%	98.19%	0.16%	NA
H3K36me3	0.46%	99.45%	0.09%	NA

【 0 5 7 7 】

30

図 1 3 は、標的タンパク質に関連するリードのピルアップを示すリードカバレッジのグラフ（図 1 1 に示す）、および遺伝子アノテーションのグラフにわたって提示されるリード対の接触マップを示す。H 3 K 4 m e 3 修飾は、近くの遺伝子の転写に関連する。密な接触（1 3 0 1）領域は、リードカバレッジの H 3 K 4 m e 3 関連ピーク（1 3 0 2）から始まり、左方向に続いているのが分かり、下に示されるように、この領域（1 3 0 3）は、同じ方向に読み取られる遺伝子を含むものとしても注釈付けされる。同様に、（1 3 0 4）において、密な接触領域が、リードカバレッジの別の H 3 K 4 m e 3 関連ピーク（1 3 0 5）から始まり、右方向に続いているのが分かり、下に示されるように、この領域（1 3 0 6）は、同じ方向に読み取られる F A B P 5 遺伝子を含むものとしても注釈付けされる。C T C F は、クロマチンループの形成に関連し、異なる遺伝子座に結合した 2 つの C T C F タンパク質が一体となり、それらの間の D N A がループを形成する。トポロジー関連ドメイン（T A D）を表す接触密度の三角形、例えば、ピーク（1 3 1 1）、（1 3 1 4）、（1 3 1 7）、および（1 3 2 0）を有するものが見られる。これらの三角形の左端および右端をなぞると、これらの領域の境界が C T C F 関連ピークと合い（例えば、（1 3 1 1）は C T C F ピーク（1 3 1 2 および 1 3 1 3）と、（1 3 1 4）は C T C F ピーク（1 3 1 5 および 1 3 1 6）と、（1 3 1 7）は C T C F ピーク（1 3 1 8 および 1 3 1 9）と、（1 3 2 0）は C T C F ピーク（1 3 2 1 および 1 3 2 2）と合う）、どの特定の C T C F 部位が一体となってどのループおよびドメインを形成するかの識別が可能となることが分かる。

40

【 0 5 7 8 】

50

図 1 4 は図 1 1 と同じ E N C O D E ピークに対する M N a s e H i C h I P 結果の比較であるが、同じ日およびその後の日の試料反復についての比較を示す。これは、プロトコルの整合性および再現性を示す。

【 0 5 7 9 】

全体として、これらの実験は、M N a s e H i C h I P ライブラリーが C h I P - s e q 特徴、M N a s e 特性、H i - C 特性を有し、タンパク質ピーク間で H i - C 相互作用を示し、プロトコルがロバストであり、高い再現性を有することを実証している。

【 0 5 8 0 】

実施例 1 2 : 分離 - プール 標識化アプローチを使用する近接ライゲーション

【 0 5 8 1 】

クロスリンキング剤で固定化された細胞を含む安定化された生体試料が得られる。試料を D N a s e で処理して、インサイチュで細胞中の D N A を消化する。その後、試料を酵素で処理して D N A の末端を平滑化し、D N A 末端をポリアデニル化する。次いで、細胞を 9 6 ウェルプレートのウェルに、ウェル当たり 1 つの細胞で分注する。バーコードを各ウェルに加え、各ウェル中の D N A 末端にライゲーションする。次いで、細胞をプールし、各ウェルに 1 つの細胞で再度分注する。第 2 のバーコードを付加し、各ウェル中の第 1 のバーコードにライゲーションする。次いで、細胞を再度プールし、ウェル当たり 1 つの細胞で再度分注し、架橋アダプターを、第 2 の架橋アダプターに適合性であるオーバーハングとライゲーションする。このアプローチを図 1 6 に示す。次いで、近位端をライゲーションして、図 1 7 に示す分子を得て、ここで、各末端は 2 つのバーコードと、他の末端上の別の架橋に結合する架橋とを有する。図 1 8 は、分割およびプールのアプローチから生じるバーコードおよび架橋の組み合わせの例を示す。次いで、クロスリンキングを脱クロスリンクさせ、核酸を精製し、配列決定して配列情報を得る。

【 0 5 8 2 】

実施例 1 3 : 標的化アダプターを使用する近接ライゲーション

【 0 5 8 3 】

クロスリンキング剤で固定化された細胞を含む安定化された生体試料が得られる。試料を D N a s e で処理して、インサイチュで細胞中の D N A を消化する。その後、試料を酵素で処理して D N A の末端を平滑化し、D N A 末端をポリアデニル化する。次いで、試料を、D N A 中のヒストンに結合する抗体と接触させ、その後、試料を、D N A 末端をライゲーションする前に、複数のプロテイン A にテザリングされたビオチン化ライブラリーアダプターと接触させる。アダプターを、ヒストン結合抗体が結合した近位端間でライゲーションする。ストレプトアビジンを用いてビオチン化アダプターをプルダウンし、得られた精製試料上でクロスリンクを脱クロスリンクする。次いで、増幅および P C R を行って、配列情報を得る。

【 0 5 8 4 】

実施例 1 4 : 凝集体三次元核酸配置の決定は、細胞特異的情報を失う可能性がある。

【 0 5 8 5 】

三次元核酸配置分析のために細胞集団を収集する。集団の細胞は、1 番染色体および 2 番染色体の領域を近接させる核酸配置を共有するが、3 番染色体および 4 番染色体の領域は、細胞のすべてではないがいくつかにおいて近接している。安定化された核を分割し、断片化して内部末端を露出させ、a t t B タグを付け、次いで、p h i C 3 1 インテグラーゼの存在下で細胞識別情報を欠く a t t P 結合核酸集団と接触させる。

【 0 5 8 6 】

ライブラリー成分を末端配列決定する。共通の分子上の 1 番染色体および 2 番染色体にマッピングされるリード対は、バックグラウンド上で示差的に観察されることが分かる。共通の分子上の 3 番染色体および 4 番染色体にマッピングされるリード対は、より低い頻度でバックグラウンド上で示差的に観察されることが分かる。

【 0 5 8 7 】

3 番染色体および 4 番染色体が近接しているが、1 番染色体および 2 番染色体よりも互

10

20

30

40

50

いに遠く離れているかどうか、または細胞集団のメンバー間で配置の変動があるかどうかを区別することはできない。

【0588】

実施例15：細胞特異的な3次元核酸配置情報は、本明細書の方法によって保存される。

【0589】

3次元核酸配置分析のために細胞集団を収集する。集団の細胞は、1番染色体および2番染色体の領域を近接させる核酸配置を共有するが、3番染色体および4番染色体の領域は、細胞のすべてではないがいくつかにおいて近接している。安定化された核を分割し、断片化して内部末端を露出させ、attBタグを付け、次いで、phiC31インテグラーゼの存在下で細胞識別情報を有するattP結合核酸の集団と接触させる。

【0590】

ライブラリー成分は、内部末端に隣接する配列、およびリンカーのパーティション識別配列の両方を得るために、末端配列決定される。共通の分子上の1番染色体および2番染色体にマッピングされるリード対は、パーティション識別配列とは無関係に、バックグラウンドの上で示差的に観察されることが分かる。共通の分子上の3番染色体および4番染色体にマッピングされるリード対は、細胞識別ライブラリー成分の第1の集団において1番染色体および2番染色体の対に匹敵するレベルで観察されるが、細胞識別ライブラリー成分の第2の集団においてバックグラウンド上では観察されないことが分かる。3番染色体および4番染色体の関連セグメントは、集団内で立体構造変化を示し、その結果、いくつかの細胞は、1番染色体および2番染色体のセグメントに匹敵する3番染色体および4番染色体のセグメントの3次元近接性を示すが、他の細胞は、1番染色体および2番染色体のセグメントに匹敵する3番染色体および4番染色体のセグメントの3次元近接性を示さないと結論付けられる。

【0591】

実施例16：細胞特異的な3次元核酸配置情報は定量的に測定される。

【0592】

3次元核酸配置分析のために細胞集団を収集する。集団の細胞は、1番染色体および2番染色体の領域を近接させる核酸配置を共有するが、3番染色体および4番染色体の領域は、細胞間でそれらの近接性が定量的に変化する。安定化された核を分割し、断片化して内部末端を露出させ、attBタグを付け、次いで、phiC31インテグラーゼの存在下で細胞識別情報を有するattP結合核酸の集団と接触させる。

【0593】

ライブラリー成分は、内部末端に隣接する配列、およびリンカーのパーティション識別配列の両方を得るために、末端配列決定される。共通の分子上の1番染色体および2番染色体にマッピングされるリード対は、パーティション識別配列とは無関係に、バックグラウンド上で示差的に観察されることが分かる。共通分子上の3番染色体および4番染色体にマッピングされるリード対は、リード対頻度と組み合わせたパーティション識別配列情報によって示されるように、細胞に応じて変化するレベルで観察されることが分かる。1番染色体および2番染色体の示されたセグメントは、細胞集団全体にわたって近接していると結論付けられる。3番染色体および4番染色体の関連セグメントは、集団内で立体構造変化を示し、その結果、3番染色体および4番染色体の示されたセグメントの近接性に関して連続体にわたって定量的変化が存在すると結論付けられる。

【0594】

実施例17：ソラレン dendrimer を用いた近接ライゲーション法

【0595】

近接ライゲーションを用いる配列決定のために細胞試料を調製する。試料中の細胞を、ホルムアルデヒドおよびジスクシンイミジルグルタレートを用いてクロスリンクする。細胞からの核を、トリトンX溶液での処理によって単離する。PAMAM dendrimer と混合された NHS-エステル-ソラレンを混合して、中性 pH で反応してソラレン分子をデ

ンドリマーにコンジュゲートさせる1級アミンで各ポリマーを終結させることによって作製されたソラレンデンドリマーを核に添加し、360 nmの光で光活性化する。核をDNaseで消化し、末端をT/Aライゲーションのために固定化する。アダプターを末端にライゲーションする。余剰なアダプターを洗い流す。アダプターは、近接ライゲーションを行う前にリン酸化される。ソラレンクロスリンクは、熱アルカリ処理または254 nmでのUV照射によって脱クロスリンクされる。DNAを混合物から精製し、配列決定する。

【0596】

本方法は、コンカテマーを生成し、より多くのクロスリンク事象を有する閉じ込め領域を有する空間内に複数のDNA断片を維持し、それによって複数の接合部を形成する。

10

【0597】

実施例18：循環を用いる近接ライゲーション方法

【0598】

核酸試料を断片化し、Tn5トランスポザーゼを用いてattBでタグ付けした。核酸を捕捉ビーズに結合させ、Phic31を使用して、隣接するタグ付き断片を結合し、架橋オリゴヌクレオチドアダプターによって結合された2つの断片を有する組換え産物を作製した。核酸単離の前に、組換え産物の末端を除去し、クロスリンクを脱クロスリンクさせた。単離された組換え産物を環状化し、PCRを用いて試料核酸を増幅した。PCR産物を精製し、配列決定前にサイズ選択に供する(図32および図33)。この方法は、改善されたロングレンジ情報をもたらす(図38)。

20

【0599】

実施例19：核試料調製

【0600】

62℃で0.3%SDSで細胞を処理することによって、近接ライゲーションのために核を調製した。核を遠心分離によって単離した後、上記のタグメンテーションおよび組換えの工程を行った。試料調製物は、改善されたカバレッジ均一性(図34)、同等の有効なリードパーセンテージ、cis>1kbのパーセンテージ、cis>10kbのパーセンテージ、cis>1Mbのパーセンテージ、および400Mでの複雑性(図35および図36)を示した。調整された試料調製条件は、改善されたカバレッジおよび保持されたクロマチン立体構造情報をもたらした。

30

【0601】

実施例20：長い非コードRNA結合部位の分析

【0602】

生体試料をクロスリンクし、クロマチンを調製し、RNaseHで処理して、リボソームRNAの試料を枯渇させる。クロマチンは、切断された末端にアデニル化/ビオチン化したオリゴヌクレオチドも付加するTn5トランスポザーゼを用いて断片化される。クロマチンに結合したRNAを、T4RNAリガーゼを用いてアデニル化アダプターにライゲーションし、試料をプロテイナーゼKで処理し、クロスリンクを脱クロスリンクする。ライゲーションしたRNAの第2の鎖を逆転写酵素で伸長し、第2の鎖が産生され、DNAが精製される。ストレプトアビジンタグ付けエンドヌクレアーゼは、ビオチンタグ付けオリゴヌクレオチド付近のDNAを消化する断片に結合される。配列決定ライブラリーを調製し、ビーズを用いてビオチンタグを有するDNAを精製し、cDNA、アダプター、および結合DNAを有するライブラリーが得られる。この方法を図39に例示する。

40

【0603】

実施例21：配列決定アダプターの組み込み方法

【0604】

配列決定アダプターをトランスポザーゼに組み込むために様々な方法が使用される(図37)。図37の上部パネルにおいて、Tn5トランスポザーゼ、ならびにTn5および配列決定アダプターによって認識されるモザイク末端(ME)を含むトランスポゾン末端

50

は、トランスポソーム複合体を形成する。複合体は二本鎖DNAを断片化し、トランスポゾン二本鎖DNAの末端に結合させ、それによってDNAを断片化し、配列決定アダプターを付加する。図37の左下のパネルにおいて、モザイク末端部位、配列決定アダプター、およびA t t B部位は、トランスポザーゼとの単一反応においてトランスポゾンとしてDNAに付加される。図37の右下のパネルにおいて、モザイク末端部位および配列決定アダプターは、上のパネルのようにトランスポゾンとしてDNAに付加され、A t t B部位は、別個のライゲーション反応において付加される。第1の方法では、断片サイズは250～2500bpの範囲であり、リードの約86%はモザイク末端を有する。第2の方法では、断片サイズは25～10,000bpの範囲であり、リードの29%はモザイク末端を含み、リードの13%が2つのモザイク末端を有する。第3の方法では、断片サイズは250～2500bpの範囲であり、リードの91%はモザイク末端を含有し、リードの38%はライゲーションされたA t t Bを含有する。

10

【0605】

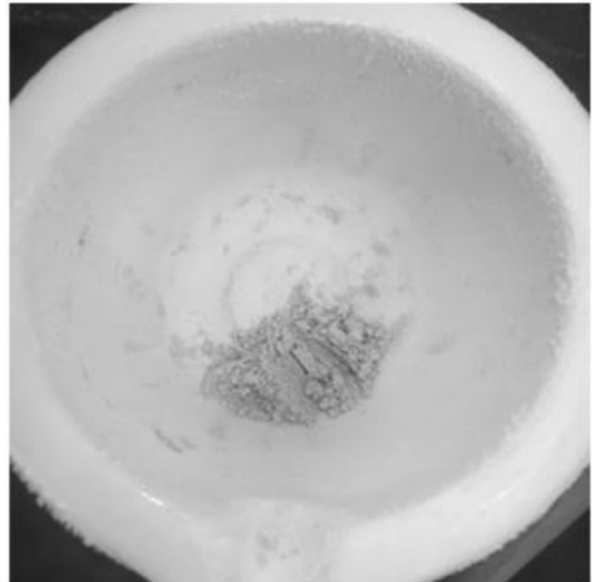
本発明の好ましい実施形態が本明細書に示され、説明されてきたが、そのような実施形態は例としてのみ提供されることが当業者には明らかであろう。当業者であれば、多くの変形、変更、および置換が、本発明から逸脱することなく思い浮かぶであろう。本明細書に記載される実施形態に対する様々な代替物が採用され得ることを理解されたい。以下の特許請求の範囲は本発明の範囲を定義し、この特許請求の範囲内の方法および構造体、ならびにその均等物がそれによって包含されることが意図されている。

20

【図面】

【図1A】

【図1B】

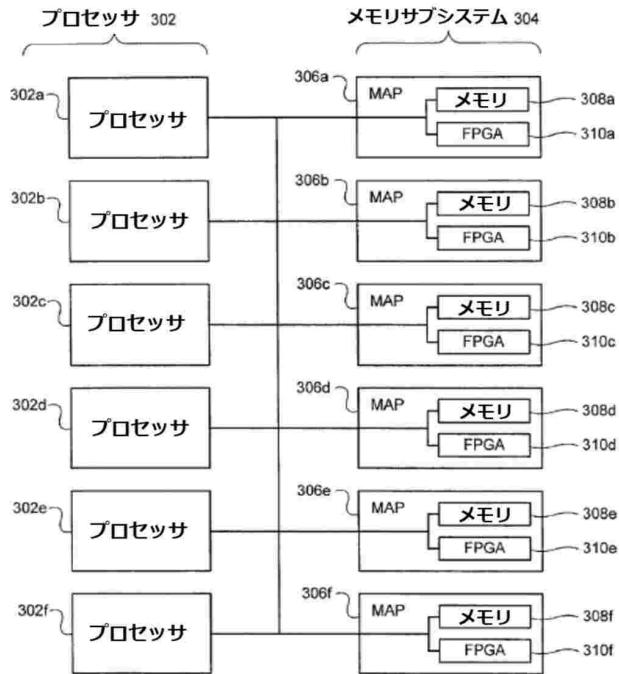


30

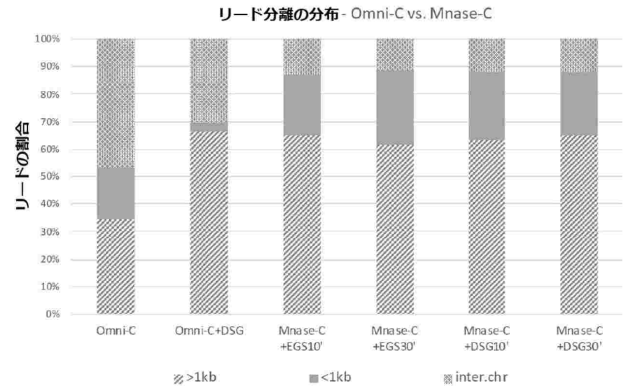
40

50

【図 6】



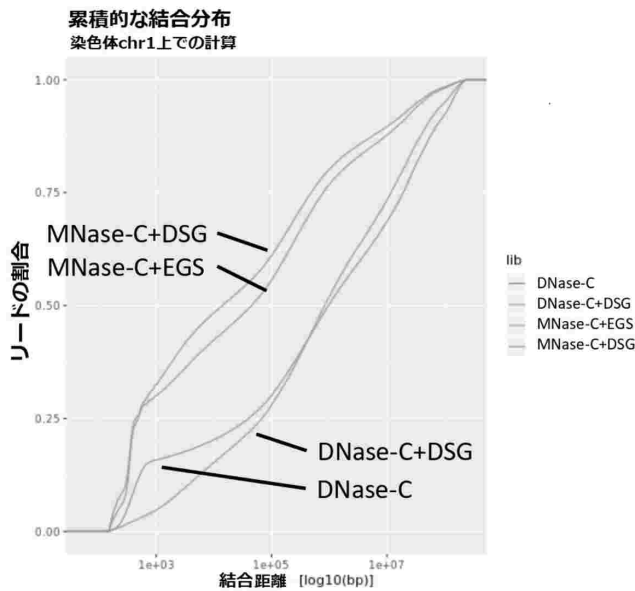
【図 7】



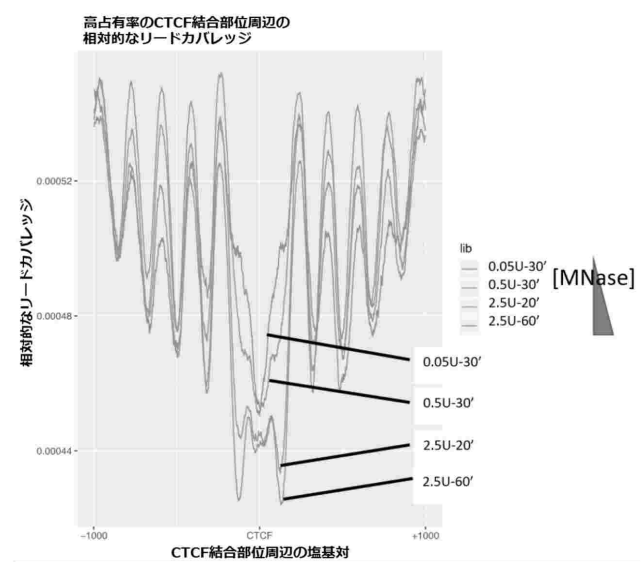
10

20

【図 8】



【図 9】

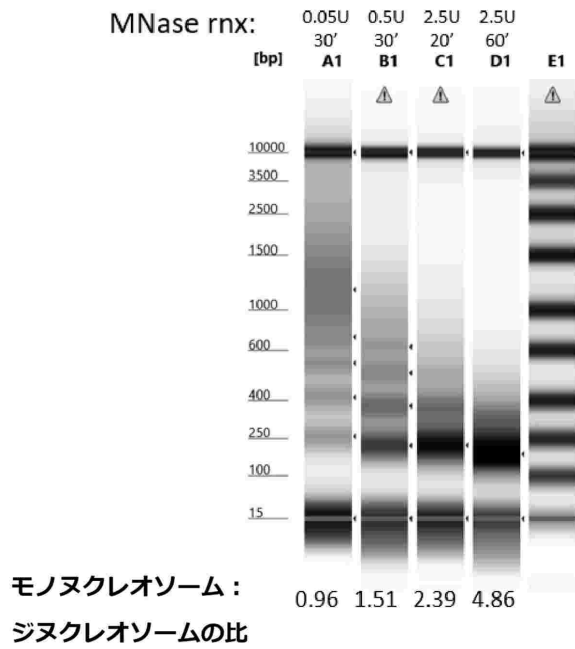


30

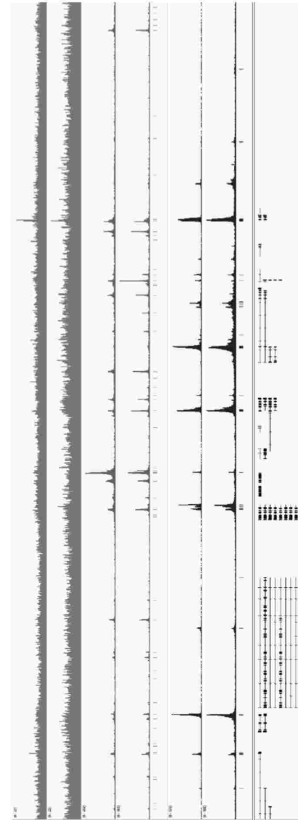
40

50

【図 10】



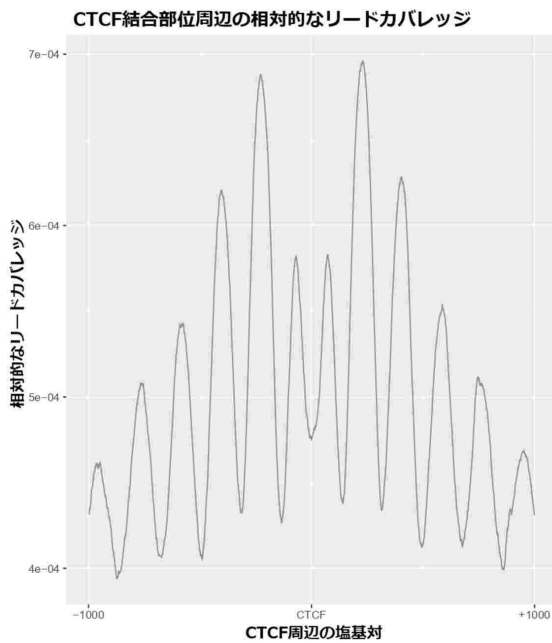
【図 11】



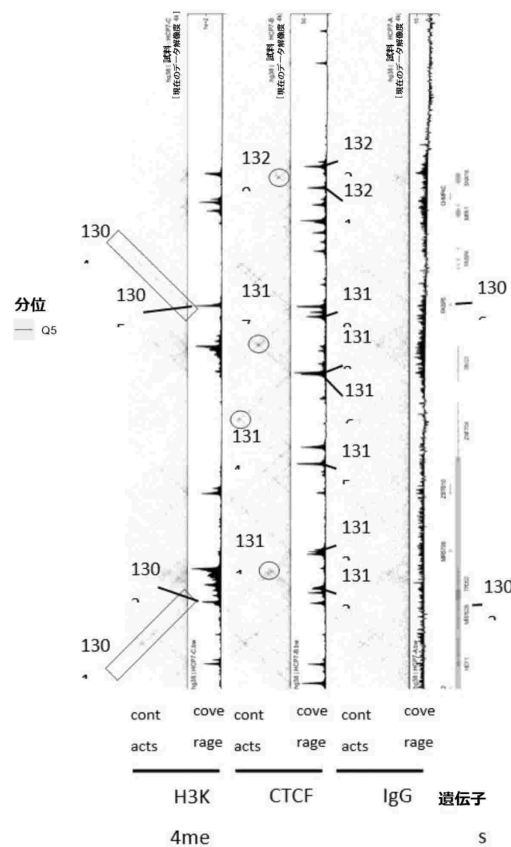
10

20

【図 12】



【図 13】

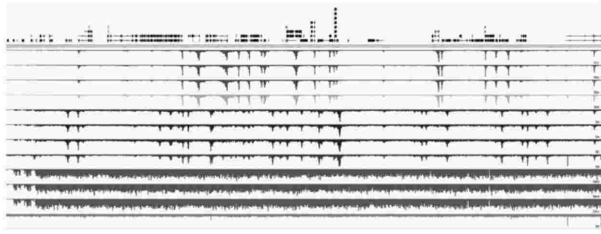


30

40

50

【図 1 4】



rep1	1	2	3	1	2	3	1	2	3	Gen
Day	2	1	2	1	2	1	2	1	2	Gen
	Neg.	CTCF			H3K					4me
	Cntl.									

【図 1 5】

インサイチュDNase消化



末端平滑化/Aデーリング



架橋アダプターライゲーション



末端の近接ライゲーション



脱クロスリンク

...NNNNNN NNNNNN...
 ...NNNNNN NNNNNN...
 ...NNNNNN A NNNNNN...
 ...NNNNNN N NNNNNN...
 ...NNNNNNNN₁₀NGA_{TC} N⁰¹NNLNNNNNN...
 ...NNNNNN TNN₁₀N N⁰¹NNNNNNNN...
 ...NNNNNNNNNN₁₀NGA_{TC} N⁰¹NNLNNNNNN...
 ...NNNNNN TNN₁₀N N⁰¹NNNNNNNN...

10

【図 1 6】

インサイチュDNase消化



末端平滑化/Aデーリング



単一細胞を分注

バーコードアダプターライゲーション



プール

単一細胞を分注

バーコードアダプターライゲーション



プール

単一細胞を分注

架橋アダプターライゲーション

...NNNNNN NNNNNN...
 ...NNNNNN NNNNNN...
 ...NNNNNN A NNNNNN...
 ...NNNNNN N NNNNNN...
 ...NNNNNNATG (N3) TCAA
 ...NNNNNN TAC (N3) A_{TC}
 ...NNNNNNATG (N3) TCAA A (N3) TCATG
 ...NNNNNN TAC (N3) AGTTT (N3) AG_{TC}
 bc1 bc2 生物学的架橋
 ...NNNNNNATG (N3) TCAA A (N3) TCATG C (N3) ATGATC
 ...NNNNNN TAC (N3) AGTTT (N3) AG TACG (N3) TA

【図 1 7】



末端の近接ライゲーション



bc1 bc2 生物学的架橋 bc2 bc1

...NNNNNNATG (N3) TCAA A (N3) TCATG C (N3) ATGATC VI (EN) SOVLGV (EN) LLLGV (EN) OVLNNNNNN...
 ...NNNNNNTAC (N3) AGTTT (N3) AG TACG (N3) TA_{TC} VGV (EN) OVLGV (EN) V VVGV (EN) OLVNNNNNN...



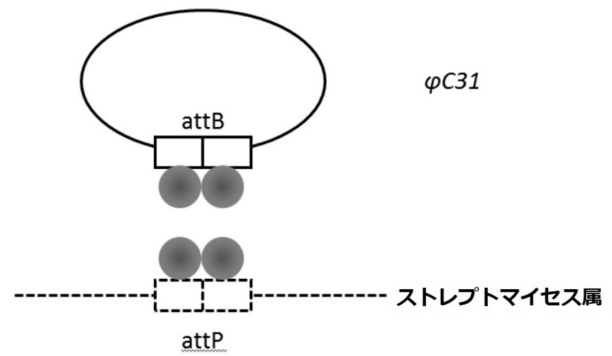
脱クロスリンク

20

【図 1 8】

bc1 bc2 生物学的架橋
 TG (N3) TCAA A (N3) TCATG C (N3) ATGATC
 TAC (N3) AGTTT (N3) AG TACG (N3) TA

【図 1 9 A】

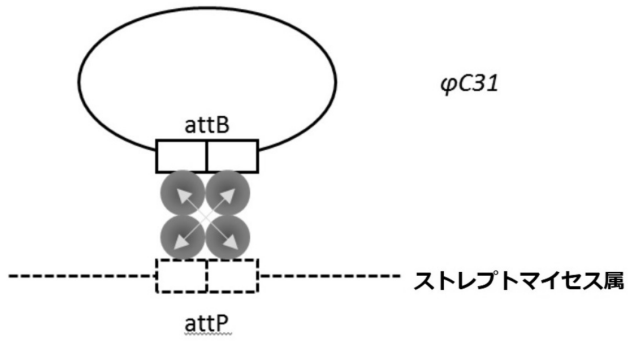


30

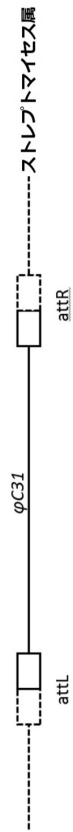
40

50

【図 19 B】



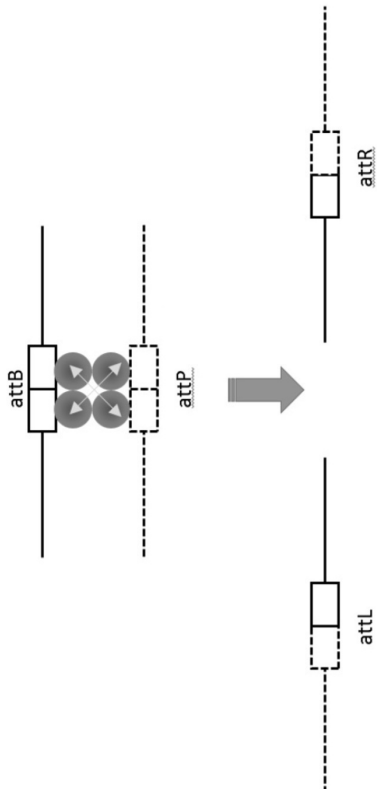
【図 19 C】



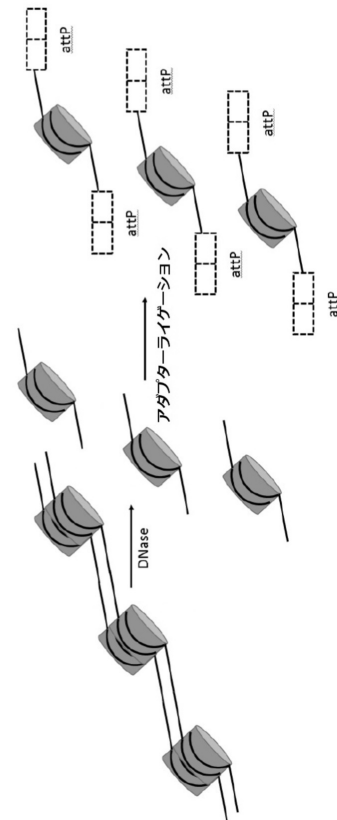
10

20

【図 19 D】



【図 20】

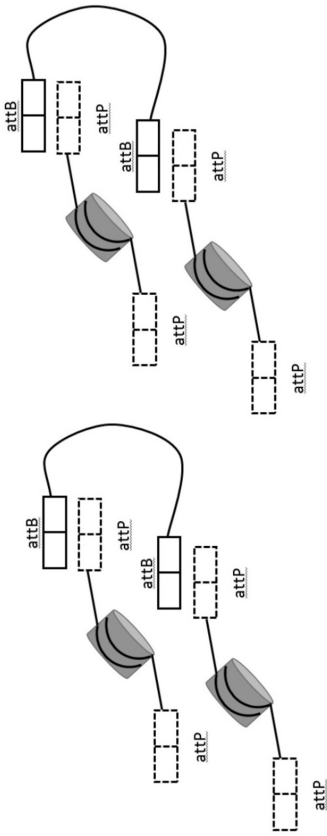


30

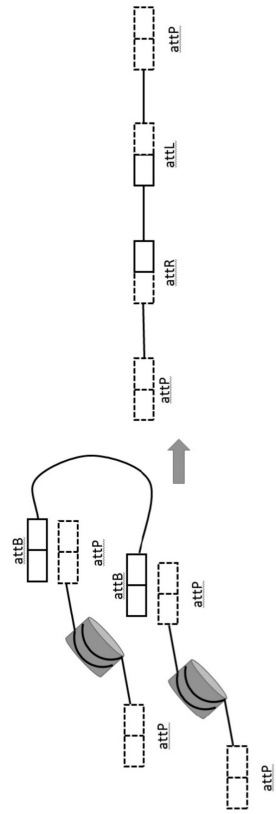
40

50

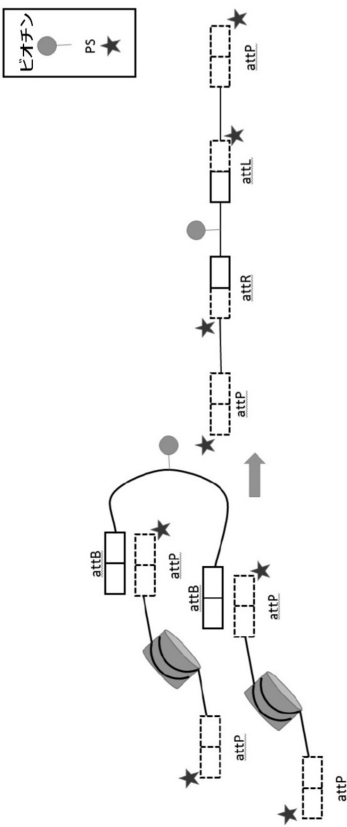
【図 2 1】



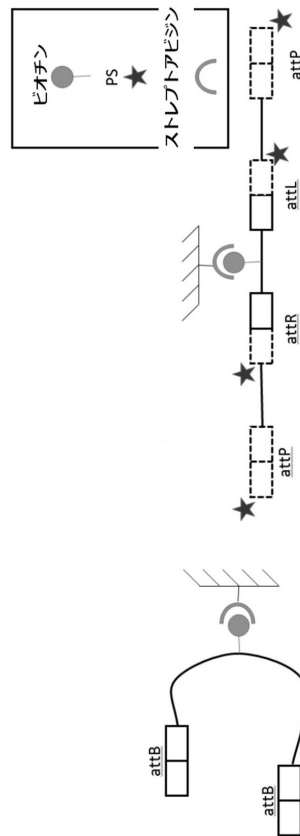
【図 2 2】



【図 2 3 A】



【図 2 3 B】



10

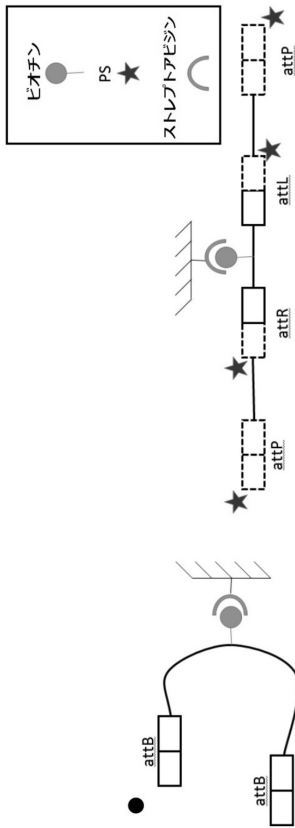
20

30

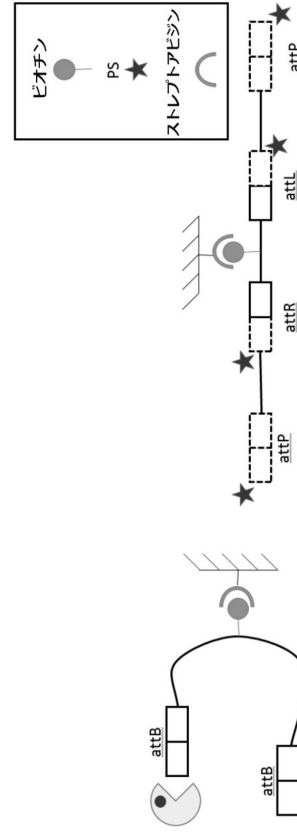
40

50

【図 2 3 C】



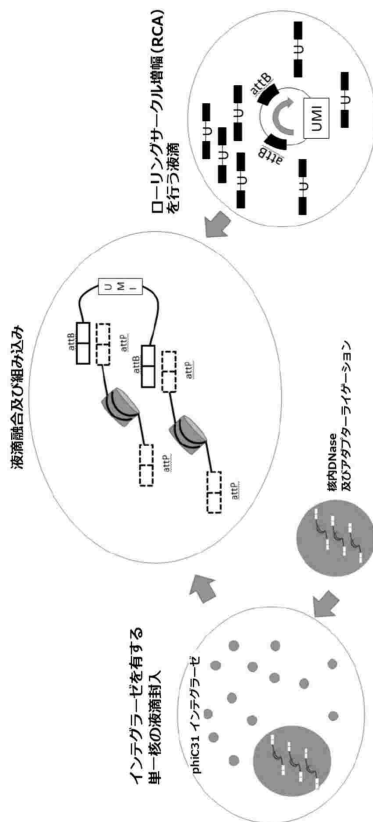
【図 2 3 D】



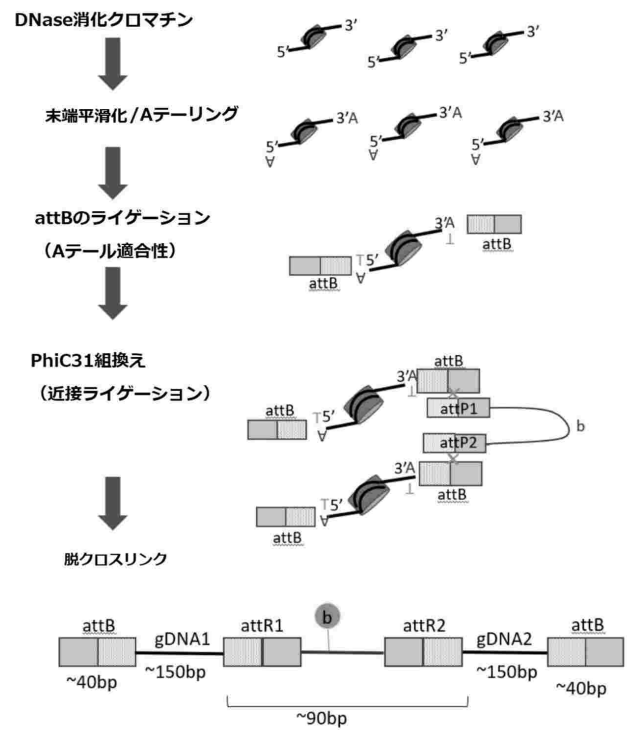
10

20

【図 2 4】



【図 2 5】



30

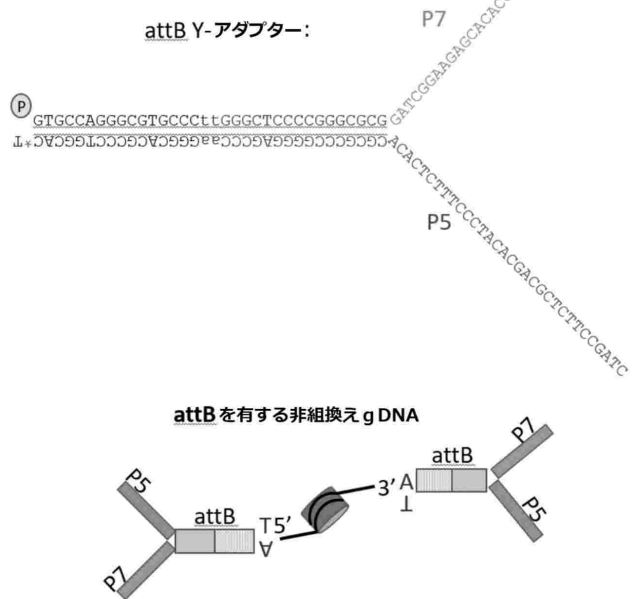
40

50

【図 26】



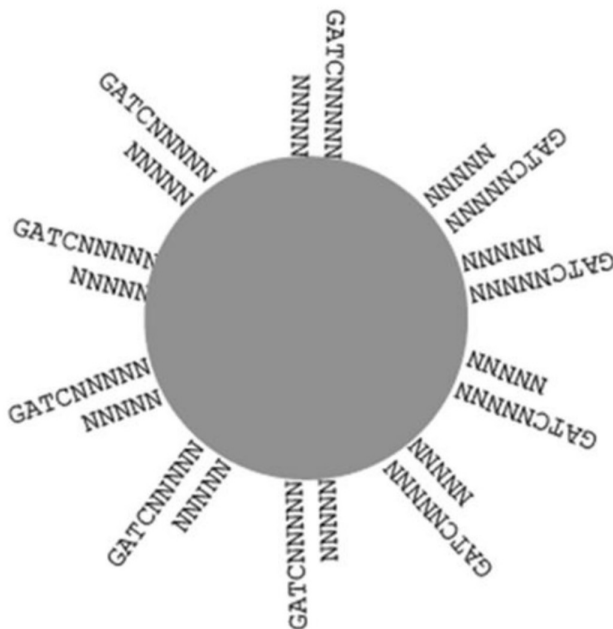
【図 27】



10

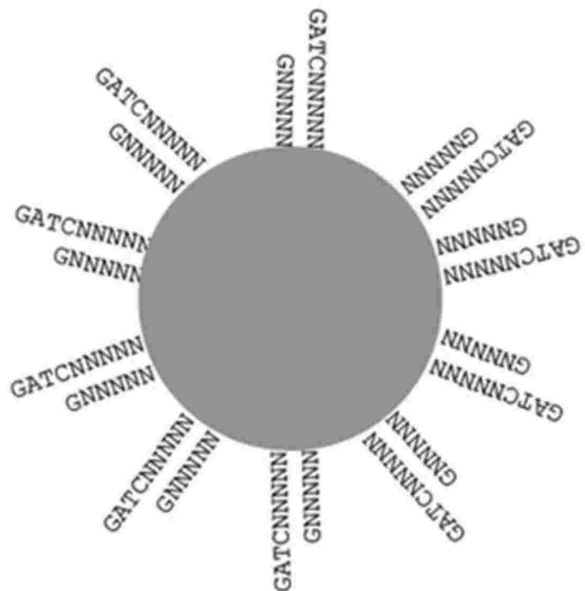
20

【図 28】



消化クロマチン凝集体

【図 29】



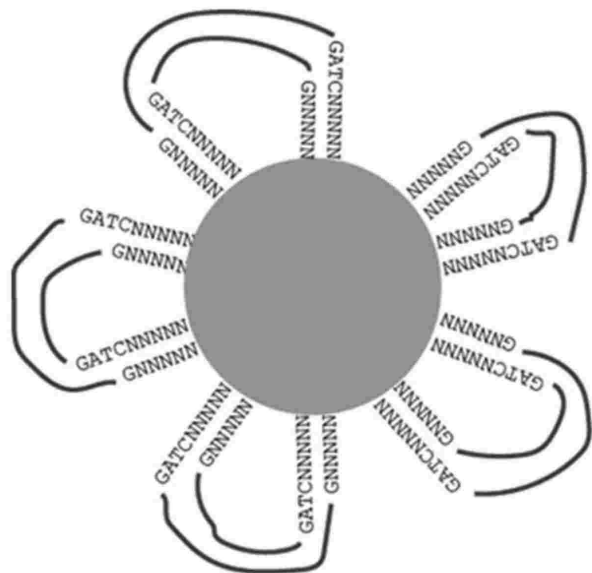
再アニーリングもライゲーションもできない
オーバーハングを作るために 1つの塩基をFill-inする

30

40

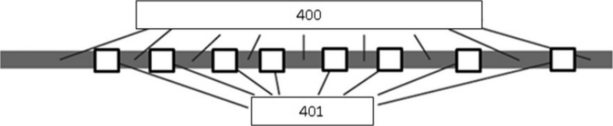
50

【 図 3 0 】

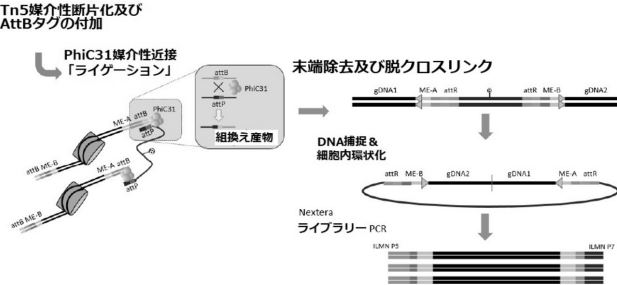


ライゲーションは切断オリゴを介して遊離末端を結合する

【 図 3 1 】



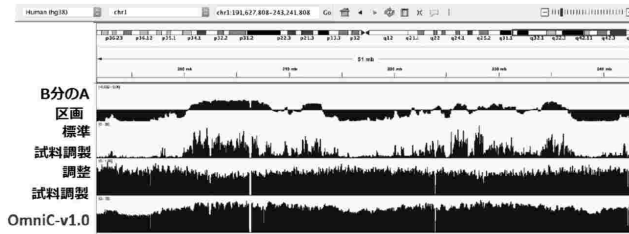
【 図 3 2 】



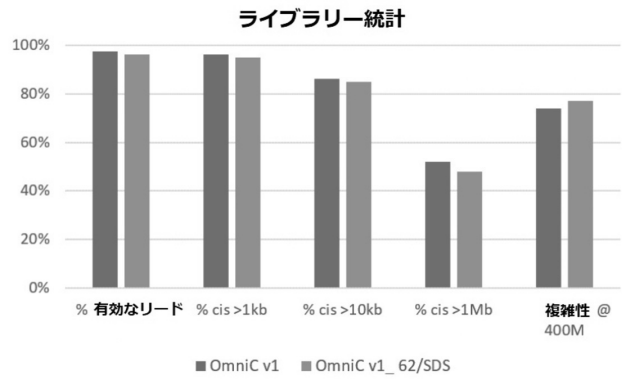
【 図 3 3 】



【図 3 4】

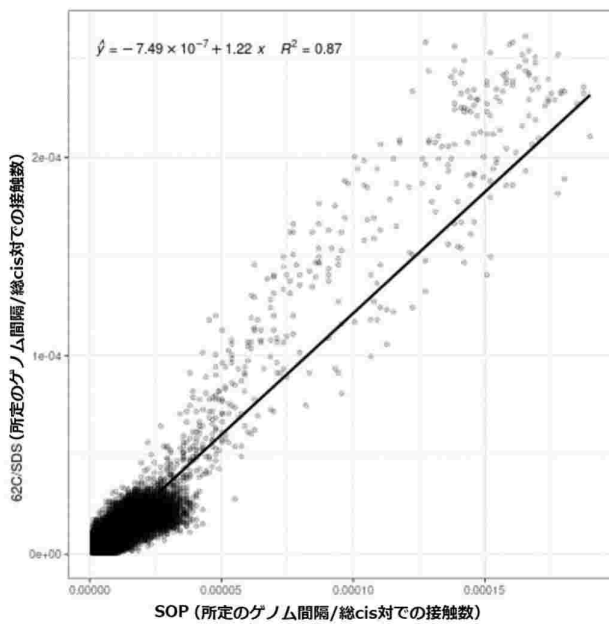


【図 3 5】

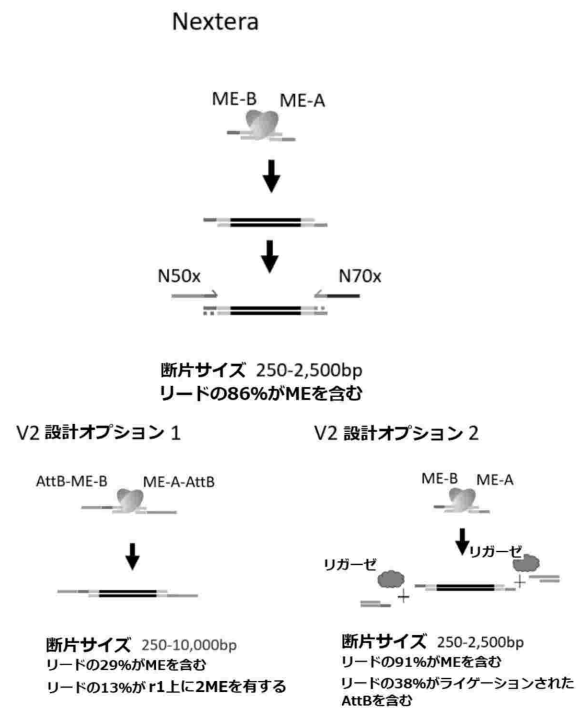


10

【図 3 6】



【図 3 7】



20

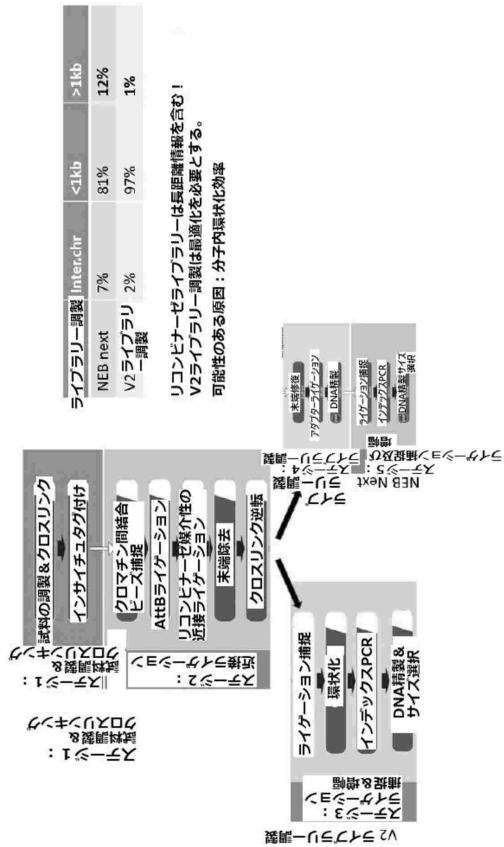
30

タグ付け後、AttBのライゲーションにより、小さな断片とより多くのattBタグ付けした末端が生成される

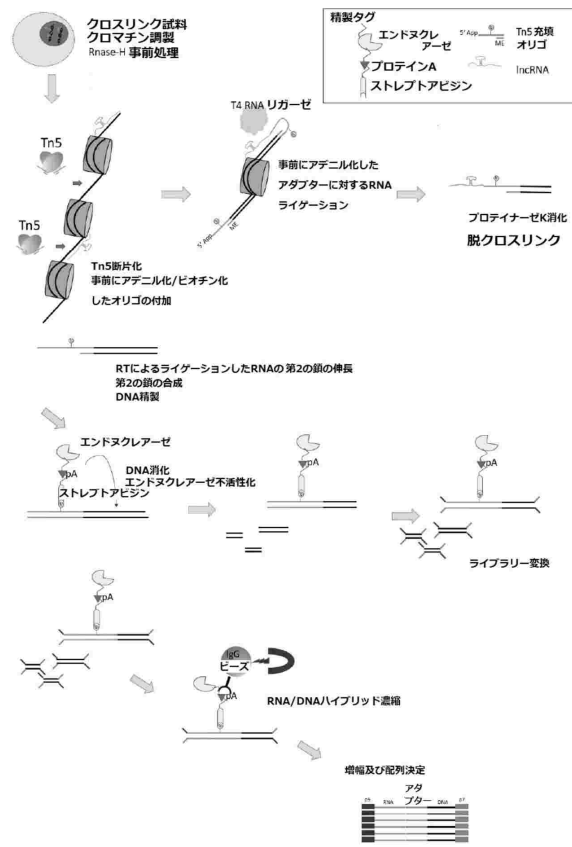
40

50

【図 38】



【図 39】



10

20

【配列表】

2024502028000001.app

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/065519

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/22; C12N 15/10; C12P 19/34; C12Q 1/6806 (2022.01)

CPC - C12N 9/22; C12P 19/34; C12Q 1/6806; C12Q 2563/185 (2022.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2020/0370096 A1 (DOVETAIL GENOMICS LLC) 26 November 2020 (26.11.2020) entire document	1, 3-19, 21, 22, 27-35, 37-52, 54, 55, 60-66
---		---
Y		2, 20, 23-26, 36, 53, 56-59, 67-71, 119-131
Y	US 2018/0080021 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 22 March 2018 (22.03.2018) entire document	2, 36
Y	WO 2020/047002 A1 (10X GENOMICS INC.) 05 March 2020 (05.03.2020) entire document	20, 53, 119-131
Y	WO 2020/223539 A1 (THE BROAD INSTITUTE INC. et al) 05 November 2020 (05.11.2020) entire document	23-26, 56-59
Y	CORN, "PROGRAMMABLE SELF-ASSEMBLY OF DNADENDRIMER AND DNA-FULLERENE NANOSTRUCTURES," Air Force Research Laboratory, 31 October 2004 [retrieved on 15 March 2022]. Retrieved from the Internet: <https://apps.dtic.mil/sti/pdfs/ADA428157.pdf>	67-71
P, X	WO 2020/264185 A1 (DOVETAIL GENOMICS LLC) 30 December 2020 (30.12.2020) entire document	1-71, 119-131

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

18 March 2022

Date of mailing of the international search report

MAR 30 2022

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Harry Kim

Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/065519

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☐ forming part of the international application as filed:

☐ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☒ furnished subsequent to the international filing date for the purposes of international search only:

☒ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/065519

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 72-118, 132
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

フロントページの続き

(32)優先日 令和3年6月29日(2021.6.29)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 63/217,009

(32)優先日 令和3年6月30日(2021.6.30)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 63/224,722

(32)優先日 令和3年7月22日(2021.7.22)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 63/281,499

(32)優先日 令和3年11月19日(2021.11.19)

(33)優先権主張国・地域又は機関
米国(US)

(81)指定国・地域 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,T,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

Fターム(参考) 4B063 QA13 QQ42 QQ52
4B065 AA90X AA90Y CA23 CA60